(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 November 2001 (15.11.2001)

PCT

(10) International Publication Number
WO 01/85956 A2

(51) International Patent Classification[7]: C12N 15/55, 9/02, 9/20, C12Q 1/26, 1/34, 1/68, C07K 16/40, G01N 33/50, A61K 38/00

(21) International Application Number: PCT/US01/15210

(22) International Filing Date: 11 May 2001 (11.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/203,511	11 May 2000 (11.05.2000)	US
60/207,903	25 May 2000 (25.05.2000)	US
60/210,150	7 June 2000 (07.06.2000)	US
60/213,392	23 June 2000 (23.06.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US Not furnished (CIP)
Filed on Not furnished

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DAS, Debopriya** [IN/US]; Apt. 3, 1179 Bonita Avenue, Mountain View, CA 94040 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinely Avenue, #3, Sunnyvale, CA 94086 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee, #5, San Francisco, CA 94107 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **KHAN, Farrah, A.** [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place, San Jose, CA 95121 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/85956 A2

(54) Title: LIPID METABOLISM ENZYMES

(57) Abstract: The invention provides human lipid metabolism enzymes (LME) and polynucleotides which identify and encode LME. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of LME.

LIPID METABOLISM ENZYMES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of lipid metabolism enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid metabolism enzymes.

BACKGROUND OF THE INVENTION

Lipids are water-insoluble, oily or greasy substances that are soluble in nonpolar solvents such as chloroform or ether. Neutral fats (triacylglycerols) serve as major fuels and energy stores. Polar lipids, such as phospholipids, sphingolipids, glycolipids, and cholesterol, are key structural components of cell membranes. (Lipid metabolism is reviewed in Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY; Lehninger, A. (1982) Principles of Biochemistry, Worth Publishers, Inc. New York NY; and ExPASy "Biochemical Pathways" index of Boehringer Mannheim World Wide Web site, "http://www.expasy.ch/cgi-bin/search-biochem-index".)

Fatty acids are long-chain organic acids with a single carboxyl group and a long non-polar hydrocarbon tail. Long-chain fatty acids are essential components of glycolipids, phospholipids, and cholesterol, which are building blocks for biological membranes, and of triglycerides, which are biological fuel molecules. Long-chain fatty acids are also substrates for eicosanoid production, and are important in the functional modification of certain complex carbohydrates and proteins. 16-carbon and 18-carbon fatty acids are the most common. Fatty acid synthesis occurs in the cytoplasm. In the first step, acetyl-Coenzyme A (CoA) carboxylase (ACC) synthesizes malonyl-CoA from acetyl-CoA and bicarbonate. The enzymes which catalyze the remaining reactions are covalently linked into a single polypeptide chain, referred to as the multifunctional enzyme fatty acid synthase (FAS). FAS catalyzes the synthesis of palmitate from acetyl-CoA and malonyl-CoA. FAS contains acetyl transferase, malonyl transferase, β-ketoacetyl synthase, acyl carrier protein, β-ketoacyl reductase, dehydratase, enoyl reductase, and thioesterase activities. The final product of the FAS reaction is the 16-carbon fatty acid palmitate. Further elongation, as well as unsaturation, of palmitate by accessory enzymes of the ER produces the variety of long chain fatty acids required by the individual cell. These enzymes include a NADH-cytochrome b_5 reductase, cytochrome b_5, and a desaturase.

Triacylglycerols, also known as triglycerides and neutral fats, are major energy stores in animals. Triacylglycerols are esters of glycerol with three fatty acid chains. Glycerol-3-phosphate is produced from dihydroxyacetone phosphate by the enzyme glycerol phosphate dehydrogenase or

from glycerol by glycerol kinase. Fatty acid-CoA's are produced from fatty acids by fatty acyl-CoA synthetases. Glyercol-3-phosphate is acylated with two fatty acyl-CoA's by the enzyme glycerol phosphate acyltransferase to give phosphatidate. Phosphatidate phosphatase converts phosphatidate to diacylglycerol, which is subsequently acylated to a triacylglyercol by the enzyme diglyceride acyltransferase. Phosphatidate phosphatase and diglyceride acyltransferase form a triacylglyerol synthetase complex bound to the ER membrane.

A major class of phospholipids are the phosphoglycerides, which are composed of a glycerol backbone, two fatty acid chains, and a phosphorylated alcohol. Phosphoglycerides are components of cell membranes. Principal phosphoglycerides are phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl serine, phosphatidyl inositol, and diphosphatidyl glycerol. Many enzymes involved in phosphoglyceride synthesis are associated with membranes (Meyers, R.A. (1995) Molecular Biology and Biotechnology, VCH Publishers Inc., New York NY, pp. 494-501). Phosphatidate is converted to CDP-diacylglycerol by the enzyme phosphatidate cytidylyltransferase (ExPASy ENZYME EC 2.7.7.41). Transfer of the diacylglycerol group from CDP-diacylglycerol to serine to yield phosphatidyl serine, or to inositol to yield phosphatidyl inositol, is catalyzed by the enzymes CDP-diacylglycerol-serine O-phosphatidyltransferase and CDP-diacylglycerol-inositol 3-phosphatidyltransferase, respectively (ExPASy ENZYME EC 2.7.8.8; ExPASy ENZYME EC 2.7.8.11). The enzyme phosphatidyl serine decarboxylase catalyzes the conversion of phosphatidyl serine to phosphatidyl ethanolamine, using a pyruvate cofactor (Voelker, D.R. (1997) Biochim. Biophys. Acta 1348:236-244). Phosphatidyl choline is formed using diet-derived choline by the reaction of CDP-choline with 1,2-diacylglycerol, catalyzed by diacylglycerol cholinephosphotransferase (ExPASy ENZYME 2.7.8.2).

Cholesterol, composed of four fused hydrocarbon rings with an alcohol at one end, moderates the fluidity of membranes in which it is incorporated. In addition, cholesterol is used in the synthesis of steroid hormones such as cortisol, progesterone, estrogen, and testosterone. Bile salts derived from cholesterol facilitate the digestion of lipids. Cholesterol in the skin forms a barrier that prevents excess water evaporation from the body. Farnesyl and geranylgeranyl groups, which are derived from cholesterol biosynthesis intermediates, are post-translationally added to signal transduction proteins such as Ras and protein-targeting proteins such as Rab. These modifications are important for the activities of these proteins (Guyton, A.C. (1991) Textbook of Medical Physiology, W.B. Saunders Company, Philadelphia PA, pp. 760-763; Stryer, supra, pp. 279-280, 691-702, 934). Mammals obtain cholesterol derived from both de novo biosynthesis and the diet.

Sphingolipids are an important class of membrane lipids that contain sphingosine, a long chain amino alcohol. They are composed of one long-chain fatty acid, one polar head alcohol, and sphingosine or sphingosine derivatives. The three classes of sphingolipids are sphingomyelins,

cerebrosides, and gangliosides. Sphingomyelins, which contain phosphocholine or phosphoethanolamine as their head group, are abundant in the myelin sheath surrounding nerve cells. Galactocerebrosides, which contain a glucose or galactose head group, are characteristic of the brain. Other cerebrosides are found in non-neural tissues. Gangliosides, whose head groups contain multiple sugar units, are abundant in the brain, but are also found in non-neural tissues.

Eicosanoids, including prostaglandins, prostacyclin, thromboxanes, and leukotrienes, are 20-carbon molecules derived from fatty acids. Eicosanoids are signaling molecules which have roles in pain, fever, and inflammation. The precursor of all eicosanoids is arachidonate, which is generated from phospholipids by phospholipase A_2 and from diacylglycerols by diacylglycerol lipase. Leukotrienes are produced from arachidonate by the action of lipoxygenases.

Within cells, fatty acids are transported by cytoplasmic fatty acid binding proteins (Online Mendelian Inheritance in Man (OMIM) *134650 Fatty Acid-Binding Protein 1, Liver; FABP1). Diazepam binding inhibitor (DBI), also known as endozepine and acyl CoA-binding protein, is an endogenous γ-aminobutyric acid (GABA) receptor ligand which is thought to down-regulate the effects of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters (OMIM *125950 Diazepam Binding Inhibitor; DBI; PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Fat stored in liver and adipose triglycerides may be released by hydrolysis and transported in the blood. Free fatty acids are transported in the blood by albumin. Triacylglycerols and cholesterol esters in the blood are transported in lipoprotein particles. The particles consist of a core of hydrophobic lipids surrounded by a shell of polar lipids and apolipoproteins. The protein components serve in the solubilization of hydrophobic lipids and also contain cell-targeting signals. Lipoproteins include chylomicrons, chylomicron remnants, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). There is a strong inverse correlation between the levels of plasma HDL and risk of premature coronary heart disease.

Mitochondrial and peroxisomal beta-oxidation enzymes degrade saturated and unsaturated fatty acids by sequential removal of two-carbon units from CoA-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids. The pathways of mitochondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) Biochem. J.

320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) Biochimie 75:147-158). Enzymes involved in beta-oxidation include acyl CoA synthetase, carnitine acyltransferase, acyl CoA dehydrogenases, enoyl CoA hydratases, L-3-hydroxyacyl CoA dehydrogenase, β-ketothiolase, 2,4-dienoyl CoA reductase, and isomerase.

Three classes of lipid metabolism enzymes are discussed in further detail. The three classes are lipases, phospholipases and lipoxygenases.

Lipases

Triglycerides are hydrolyzed to fatty acids and glycerol by lipases. Adipocytes contain lipases that break down stored triacylglycerols, releasing fatty acids for export to other tissues where they are required as fuel. Lipases are widely distributed in animals, plants, and prokaryotes. Triglyceride lipases (ExPASy ENZYME EC 3.1.1.3), also known as triacylglycerol lipases and tributyrases, hydrolyze the ester bond of triglycerides. In higher vertebrates there are at least three tissue-specific isozymes including gastric, hepatic, and pancreatic lipases. These three types of lipases are structurally closely related to each other as well as to lipoprotein lipase. The most conserved region in gastric, hepatic, and pancreatic lipases is centered around a serine residue which is also present in lipases of prokaryotic origin. Mutation in the serine residue renders the enzymes inactive. Gastric, hepatic, and pancreatic lipases hydrolyze lipoprotein triglycerides and phospholipids. Gastric lipases in the intestine aid in the digestion and absorption of dietary fats. Hepatic lipases are bound to and act at the endothelial surfaces of hepatic tissues. Hepatic lipases also play a major role in the regulation of plasma lipids. Pancreatic lipase requires a small protein cofactor, colipase, for efficient dietary lipid hydrolysis. Colipase binds to the C-terminal, non-catalytic domain of lipase, thereby stabilizing an active conformation and considerably increasing the overall hydrophobic binding site. Deficiencies of these enzymes have been identified in man, and all are associated with pathologic levels of circulating lipoprotein particles (Gargouri, Y. et al. (1989) Biochim. Biophys. Acta 1006:255-271; Connelly, P.W. (1999) Clin. Chim. Acta 286:243-255; van Tilbeurgh, H. et al. (1999) Biochim Biophys Acta 1441:173-184).

Lipoprotein lipases (ExPASy ENZYME EC 3.1.1.34), also known as clearing factor lipases, diglyceride lipases, or diacylglycerol lipases, hydrolyze triglycerides and phospholipids present in circulating plasma lipoproteins, including chylomicrons, very low and intermediate density lipoproteins, and high-density lipoproteins (HDL). Together with pancreatic and hepatic lipases, lipoprotein lipases (LPL) share a high degree of primary sequence homology. Both lipoprotein

lipases and hepatic lipases are anchored to the capillary endothelium via glycosaminoglycans and can be released by intravenous administration of heparin. LPLs are primarily synthesized by adipocytes, muscle cells, and macrophages. Catalytic activities of LPLs are activated by apolipoprotein C-II and are inhibited by high ionic strength conditions such as 1 M NaCl. LPL deficiencies in humans contribute to metabolic diseases such as hypertriglyceridemia, HDL2 deficiency, and obesity (Jackson, R.L. (1983) in The Enzymes (Boyer, P.D., ed.) Vol. XVI, pp. 141-186, Academic Press, New York NY; Eckel, R.H. (1989) New Engl. J. Med. 320:1060-1068).

Phospholipases

Phospholipases, a group of enzymes that catalyze the hydrolysis of membrane phospholipids, are classified according to the bond cleaved in a phospholipid. They are classified into PLA1, PLA2, PLB, PLC, and PLD families. Phospholipases are involved in many inflammatory reactions by making arachidonate available for eicosanoid biosynthesis. More specifically, arachidonic acid is processed into bioactive lipid mediators of inflammation such as lyso-platelet-activating factor and eicosanoids. The synthesis of arachidonic acid from membrane phospholipids is the rate-limiting step in the biosynthesis of the four major classes of eicosanoids (prostaglandins, prostacyclins, thromboxanes and leukotrienes) which are involved in pain, fever, and inflammation (Kaiser, E. et al. (1990) Clin. Biochem. 23:349-370). Furthermore, leukotriene-B4 is known to function in a feedback loop which further increases PLA2 activity (Wijkander, J. et al. (1995) J. Biol. Chem. 270:26543-26549).

The secretory phospholipase A_2 (PLA2) superfamily comprises a number of heterogeneous enzymes whose common feature is to hydrolyze the sn-2 fatty acid acyl ester bond of phosphoglycerides. Hydrolysis of the glycerophospholipids releases free fatty acids and lysophospholipids. PLA2 activity generates precursors for the biosynthesis of biologically active lipids, hydroxy fatty acids, and platelet-activating factor. PLA2s were first described as components of snake venoms, and were later characterized in numerous species. PLA2s have traditionally been classified into several major groups and subgroups based on their amino acid sequences, divalent cation requirements, and location of disulfide bonds. The PLA2s of Groups I, II, and III consist of low molecular weight, secreted, Ca^{2+}-dependent proteins. Group IV PLA2s are primarily 85-kDa, Ca^{2+}-dependent cytosolic phospholipases. Finally, a number of Ca^{2+}-independent PLA2s have been described, which comprise Group V (Davidson, F.F. and E.A. Dennis (1990) J. Mol. Evol. 31:228-238; and Dennis, E.F. (1994) J. Biol Chem. 269:13057-13060).

The first PLA2s to be extensively characterized were the Group I, II, and III PLA2s found in snake and bee venoms. These venom PLA2s share many features with mammalian PLA2s including a common catalytic mechanism, the same Ca^{2+} requirement, and conserved primary and tertiary structures. In addition to their role in the digestion of prey, the venom PLA2s display neurotoxic,

myotoxic, anticoagulant, and proinflammatory effects in mammalian tissues. This diversity of pathophysiological effects is due to the presence of specific, high affinity receptors for these enzymes on various cells and tissues (Lambeau, G. et al. (1995) J. Biol. Chem. 270:5534-5540).

PLA2s from Groups I, IIA, IIC, and V have been described in mammalian and avian cells, and were originally characterized by tissue distribution, although the distinction is no longer absolute. Thus, Group I PLA2s were found in the pancreas, Group IIA and IIC were derived from inflammation-associated tissues (e.g., the synovium), and Group V were from cardiac tissue. The pancreatic PLA2s function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. The Group II inflammatory PLA2s are potent mediators of inflammatory processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. These Group II PLA2s are found in most human cell types assayed and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia. A Group V PLA2 has been cloned from brain tissue and is strongly expressed in heart tissue. A human PLA2 was recently cloned from fetal lung, and based on its structural properties, appears to be the first member of a new group of mammalian PLA2s, referred to as Group X. Other PLA2s have been cloned from various human tissues and cell lines, suggesting a large diversity of PLA2s (Chen, J. et al. (1994) J. Biol. Chem. 269:2365-2368; Kennedy, B.P. et al. (1995) J. Biol. Chem. 270: 22378-22385; Komada, M. et al. (1990) Biochem. Biophys. Res. Commun. 168:1059-1065; Cupillard, L. et al. (1997) J. Biol. Chem. 272:15745-15752; and Nalefski, E.A. et al. (1994) J. Biol. Chem. 269:18239-18249).

Phospholipases B (PLB) (ExPASy ENZYME EC 3.1.1.5), also known as lysophospholipase, lecithinase B, or lysolecithinase are widely distributed enzymes that metabolize intracellular lipids, and occur in numerous isoforms. Small isoforms, approximately 15-30 kD, function as hydrolases; large isoforms, those exceeding 60 kD, function both as hydrolases and transacylases. A particular substrate for PLBs, lysophosphatidylcholine, causes lysis of cell membranes when it is formed or imported into a cell. PLBs are regulated by lipid factors including acylcarnitine, arachidonic acid, and phosphatidic acid. These lipid factors are signaling molecules important in numerous pathways, including the inflammatory response (Anderson, R. et al. (1994) Toxicol. Appl. Pharmacol. 125:176-183; Selle, H. et al. (1993); Eur. J. Biochem. 212:411-416).

Phospholipase C (PLC) (ExPASy ENZYME EC 3.1.4.10) plays an important role in transmembrane signal transduction. Many extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate PLCs. The role of an activated PLC is to catalyze the hydrolysis of phosphatidyl-inositol-4, 5-bisphosphate (PIP2), a minor component of the plasma membrane, to produce diacylglycerol and inositol 1, 4, 5-trisphosphate (IP3). In their respective biochemical

pathways, IP3 and diacylglycerol serve as second messengers and trigger a series of intracellular responses. IP3 induces the release of Ca^{2+} from internal cellular storage, and diacylglycerol activates protein kinase C (PKC). Both pathways are part of transmembrane signal transduction mechanisms which regulate cellular processes which include secretion, neural activity, metabolism, and proliferation.

Several distinct isoforms of PLC have been identified and are categorized as PLC-beta, PLC-gamma, and PLC-delta. Subtypes are designated by adding Arabic numbers after the Greek letters, eg. PLC-ß-1. PLCs have a molecular mass of 62-68 kDa, and their amino acid sequences show two regions of significant similarity. The first region designated X has about 170 amino acids, and the second or Y region contains about 260 amino acids.

The catalytic activities of the three isoforms of PLC are dependent upon Ca^{2+}. It has been suggested that the binding sites for Ca^{2+} in the PLCs are located in the Y-region, one of two conserved regions. The hydrolysis of common inositol-containing phospholipids, such as phosphatidylinositol (PI), phosphatidylinositol 4-monophosphate (PIP), and phosphatidylinositol 4, 5-bisphosphate (PIP2), by any of the isoforms yields cyclic and noncyclic inositol phosphates (Rhee, S.G. and Y.S. Bae (1997) J. Biol. Chem. 272:15045-15048).

All mammalian PLCs contain a pleckstrin homology (PH) domain which is about 100 amino acids in length and is composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. PH domains target PLCs to the membrane surface by interacting with either the beta/gamma subunits of G proteins or PIP2 (PROSITE PDOC50003).

Phospholipase D (PLD) (ExPASy ENZYME EC 3.1.4.4), also known as lecithinase D, lipophosphodiesterase II, and choline phosphatase catalyzes the hydrolysis of phosphatidylcholine and other phospholipids to generate phosphatidic acid. PLD plays an important role in membrane vesicle trafficking, cytoskeletal dynamics, and transmembrane signal transduction. In addition, the activation of PLD is involved in cell differentiation and growth (reviewed in Liscovitch, M. (2000) Biochem. J. 345:401-415).

PLD is activated in mammalian cells in response to diverse stimuli that include hormones, neurotransmitters, growth factors, cytokines, activators of protein kinase C, and agonists binding to G-protein-coupled receptors. At least two forms of mammalian PLD, PLD1 and PLD2, have been identified. PLD1 is activated by protein kinase C alpha and by the small GTPases ARF and RhoA. (Houle, M.G. and S. Bourgoin (1999) Biochim. Biophys. Acta 1439:135-149). PLD2 can be selectively activated by unsaturated fatty acids such as oleate (Kim, J.H. (1999) FEBS Lett. 454:42-46).

Lipoxygenases

Lipoxygenases (ExPASy ENZYME EC 1.13.11.12) are non-heme iron-containing enzymes

that catalyze the dioxygenation of certain polyunsaturated fatty acids such as lipoproteins. Lipoxygenases are found widely in plants, fungi, and animals. Several different lipoxygenase enzymes are known, each having a characteristic oxidation action. In animals, there are specific lipoxygenases that catalyze the dioxygenation of arachidonic acid at the carbon-3, 5, 8, 11, 12, and 15 positions. These enzymes are named after the position of arachidonic acid that they dioxygenate. Lipoxygenases have a single polypeptide chain with a molecular mass of ~75-80 kDa in animals. The proteins have an N-terminal-barrel domain and a larger catalytic domain containing a single atom of non-heme iron. Oxidation of the ferric enzyme to an active form is required for catalysis (Yamamoto, S. (1992) Biochim. Biophys. Acta 1128:117-131; Brash, A.R. (1999) J. Biol. Chem. 274:23679-23682). A variety of lipoxygenase inhibitors exist and are classified into five major categories according to their mechanism of inhibition. These include antioxidants, iron chelators, substrate analogues, lipoxygenase-activating protein inhibitors, and, finally, epidermal growth factor-receptor inhibitors.

3-Lipoxygenase, also known as e-LOX-3 or Aloxe3 has recently been cloned from murine epidermis. Aloxe3 resides on mouse chromosome 11, and the deduced amino acid sequence for Aloxe3 is 54% identical to the 12-lipoxygenase sequences (Kinzig, A. (1999) Genomics 58:158-164).

5-Lipoxygenase (5-LOX, ExPASy ENZYME EC 1.13.11.34), also known as arachidonate:oxygen 5-oxidoreductase, is found primarily in white blood cells, macrophages, and mast cells. 5-LOX converts arachidonic acid first to 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and then to leukotriene (LTA4 (5,6-oxido-7,9,11,14-eicosatetraenoic acid)). Subsequent conversion of leukotriene A4 by leukotriene A4 hydrolase yields the potent neutrophil chemoattractant leukotriene B4. Alternatively, conjugation of LTA4 with glutathione by leukotriene C4 synthase plus downstream metabolism leads to the cysteinyl leukotrienes that influence airway reactivity and mucus secretion, especially in asthmatics. Most lipoxygenases require no other cofactors or proteins for activity. In contrast, the mammalian 5-LOX requires calcium and ATP, and is activated in the presence of a 5-LOX activating protein (FLAP). FLAP itself binds to arachidonic acid and supplies 5-LOX with substrate (Lewis, R.A. et al. (1990) New Engl. J. Med. 323:645-655). The expression levels of 5-LOX and FLAP are found to be increased in the lungs of patients with plexogenic (primary) pulmonary hypertension (Wright, L. et al. (1998) Am. J. Respir. Crit. Care Med. 157:219-229).

12-Lipoxygenase (12-LOX, ExPASy ENZYME: EC 1.13.11.31) oxygenates arachidonic acid to form 12-hydroperoxyeicosatetraenoic acid (12-HPETE). Mammalian 12-lipoxygenases are named after the prototypical tissues of their occurrence (hence, the leukocyte, platelet, or epidermal types). Platelet-type 12-LOX has been found to be the predominant isoform in epidermal skin specimens and epidermoid cells. Leukocyte 12-LOX was first characterized extensively from porcine leukocytes

and was found to have a rather broad distribution in mammalian tissues by immunochemical assays. Besides tissue distribution, the leukocyte 12-LOX is distinguished from the platelet-type enzyme by its ability to form 15-HPETE, in addition to 12-HPETE from arachidonic acid substrate. Leukocyte 12-LOX is highly related to 15-lipoxgenase (15-LOX) in that both are dual specificity lipoxygenases, and they are about 85% identical in primary structure in higher mammals. Leukocyte 12-LOX is found in tracheal epithelium, leukocytes, and macrophages (Conrad, D.J. (1999) Clin. Rev. Allergy Immunol.17:71-89).

15-Lipoxygenase (15-LOX; ExPASy ENZYME: EC 1.13.11.33) is found in human reticulocytes, airway epithelium, and eosinophils. 15-LOX has been detected in atherosclerotic lesions in mammals, specifically rabbit and man. The enzyme, in addition to its role in oxidative modification of lipoproteins, is important in the inflammatory reaction in atherosclerotic lesions. 15-LOX has been shown to be induced in human monocytes by the cytokine IL-4, which is known to be implicated in the inflammatory process (Kuhn, H. and S. Borngraber (1999) Adv. Exp. Med. Biol. 447:5-28).

Disease Correlation

Lipid metabolism is involved in human diseases and disorders. In the arterial disease atherosclerosis, fatty lesions form on the inside of the arterial wall. These lesions promote the loss of arterial flexibility and the formation of blood clots (Guyton, supra). In Tay-Sachs disease, the GM_2 ganglioside (a sphingolipid) accumulates in lysosomes of the central nervous system due to a lack of the enzyme N-acetylhexosaminidase. Patients suffer nervous system degeneration leading to early death (Fauci, A.S. et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p. 2171). The Niemann-Pick diseases are caused by defects in lipid metabolism. Niemann-Pick diseases types A and B are caused by accumulation of sphingomyelin (a sphingolipid) and other lipids in the central nervous system due to a defect in the enzyme sphingomyelinase, leading to neurodegeneration and lung disease. Niemann-Pick disease type C results from a defect in cholesterol transport, leading to the accumulation of sphingomyelin and cholesterol in lysosomes and a secondary reduction in sphingomyelinase activity. Neurological symptoms such as grand mal seizures, ataxia, and loss of previously learned speech, manifest 1-2 years after birth. A mutation in the NPC protein, which contains a putative cholesterol-sensing domain, was found in a mouse model of Niemann-Pick disease type C (Fauci, supra, p. 2175; Loftus, S.K. et al. (1997) Science 277:232-235).

PLAs are implicated in a variety of disease processes. For example, PLAs are found in the pancreas, in cardiac tissue, and in inflammation-associated tissues. Pancreatic PLAs function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. Inflammatory PLAs are potent mediators of inflammatory

processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. Additionally, inflammatory PLAs are found in most human cell types and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia.

The role of PLBs in human tissues has been investigated in various research studies. Hydrolysis of lysophosphatidylcholine by PLBs causes lysis in erythrocyte membranes (Selle, supra). Similarly, Endresen, M.J. et al. (1993; Scand. J. Clin. Invest. 53:733-739) reported that the increased hydrolysis of lysophosphatidylcholine by PLB in pre-eclamptic women causes release of free fatty acids into the sera. In renal studies, PLB was shown to protect Na^+,K^+-ATPase from the cytotoxic and cytolytic effects of cyclosporin A (Anderson, supra).

Lipases, phospholipases, and lipoxygenases are thought to contribute to complex diseases, such as atherosclerosis, obesity, arthritis, asthma, and cancer, as well as to single gene defects, such as Wolman's disease and Type I hyperlipoproteinemia.

The discovery of new lipid metabolism enzymes and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid metabolism enzymes.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, lipid metabolism enzymes, referred to collectively as "LME" and individually as "LME-1," "LME-2," "LME-3," "LME-4," and "LME-5." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-5.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ

ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-5. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:6-10.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

The invention further provides an isolated polynucleotide selected from the group consisting

of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic

fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional LME, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional LME, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional LME, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide

cmoprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:6-10, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological

sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"LME" refers to the amino acid sequences of substantially purified LME obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of LME. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LME either by directly interacting with LME or by acting on components of the biological pathway in which LME participates.

An "allelic variant" is an alternative form of the gene encoding LME. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding LME include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as LME or a polypeptide with at least one functional characteristic of LME. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding LME, and improper or unexpected hybridization to allelic variants,

with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding LME. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent LME. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of LME is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of LME. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LME either by directly interacting with LME or by acting on components of the biological pathway in which LME participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind LME polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies

which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic LME, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding LME or fragments of LME may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison

WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of LME or the polynucleotide encoding LME which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:6-10 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:6-10, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:6-10 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:6-10 and the region of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-5 is encoded by a fragment of SEQ ID NO:6-10. A fragment of SEQ ID NO:1-5 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-5. For example, a fragment of SEQ ID NO:1-5 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-5. The precise length of a fragment of SEQ ID NO:1-5 and the region of SEQ ID NO:1-5 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence,

for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC

concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of LME which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of LME which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of LME. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of LME.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the

antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an LME may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of LME.

"Probe" refers to nucleic acid sequences encoding LME, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that

purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is

expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing LME, nucleic acids encoding LME, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell

type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to

another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human lipid metabolism enzymes (LME), the polynucleotides encoding LME, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and

2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are lipid metabolism enzymes. For example, SEQ ID NO:1 is 94% identical, from residue P52 to Q1052, to bovine phospholipase C (g304241) as determined by the Basic Local Alignment Search Tool (BLAST). The probability score is 0.00, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance (Table 2). In an alternative example, SEQ ID NO:2 is 67% identical, from residue M44 to residue P208, to mouse group IIF secreted phospholipase A2 (g6174881) as determined by the Basic Local Alignment Search Tool (BLAST). The probability score is 6.9e-74 (Table 2). In a further example, SEQ ID NO:3 is 87% identical, from residue M1 to residue I711, to murine lipoxygenase-3 (g5304928) as determined by the Basic Local Alignment Search Tool (BLAST). The probability score is 0.0 (Table 2). SEQ ID NO:3 also contains lipoxygenase and PLAT (Polycystin-1, Lipoxygenase, Alpha-Toxin)/LH2 (Lipoxygenase homology) domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a lipoxygenase. In yet another example, SEQ ID NO:4 is 89% identical, from residue M1 to residue L1216, to bovine phospholipase C (GenBank ID g163522) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0 (Table 2). SEQ ID NO:4 also contains phosphatidylinositol-specific phospholipase C, X, and Y domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:4 is a phospholipase C. In yet another example, SEQ ID NO:5 is 38% identical, from residue A3 to residue M489 and 40% identical from residue V624 to residue G914, to murine phospholipase C-L2 (g6705987) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability scores are 3e-88 and 3e-62 respectively. SEQ ID NO:5 also contains phosphatidylinositol-specific phospholipase X and Y domains as determined by searching for statistically significant matches in the

hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:5 is a phospholipase. The algorithms and parameters for the analysis of SEQ ID NO:1-5 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:6-10 or that distinguish between SEQ ID NO:6-10 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7412054H1 is the identification number of an Incyte cDNA sequence, and BONMTUE02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71200138V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5514658) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6983515_000028_002 is the identification number of a Genscan-predicted coding sequence, with g6983515 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL1078033_00001 represents a "stitched" sequence in which 1078033 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by

an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses LME variants. A preferred LME variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the LME amino acid sequence, and which contains at least one functional or structural characteristic of LME.

The invention also encompasses polynucleotides which encode LME. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10, which encodes LME. The polynucleotide sequences of SEQ ID NO:6-10, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding LME. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding LME. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6-10. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of LME.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding LME, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring LME, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode LME and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring LME under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding LME or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding LME and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode LME and LME derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding LME or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:6-10 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding LME may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode LME may be cloned in recombinant DNA molecules that direct expression of LME, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express LME.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter LME-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of LME, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding LME may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, LME itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis

may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of LME, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active LME, the nucleotide sequences encoding LME or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding LME. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding LME. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding LME and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding LME and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding LME. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV,

or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding LME. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding LME can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding LME into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of LME are needed, e.g. for the production of antibodies, vectors which direct high level expression of LME may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of LME. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of LME. Transcription of sequences encoding LME may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.

6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding LME may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses LME in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of LME in cell lines is preferred. For example, sequences encoding LME can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,

Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding LME is inserted within a marker gene sequence, transformed cells containing sequences encoding LME can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding LME under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding LME and that express LME may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of LME using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on LME is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding LME include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding LME, or any fragments thereof, may be cloned into a vector for

the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding LME may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode LME may be designed to contain signal sequences which direct secretion of LME through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding LME may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric LME protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of LME activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a

proteolytic cleavage site located between the LME encoding sequence and the heterologous protein sequence, so that LME may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled LME may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

LME of the present invention or fragments thereof may be used to screen for compounds that specifically bind to LME. At least one and up to a plurality of test compounds may be screened for specific binding to LME. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of LME, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which LME binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express LME, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing LME or cell membrane fractions which contain LME are then contacted with a test compound and binding, stimulation, or inhibition of activity of either LME or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with LME, either in solution or affixed to a solid support, and detecting the binding of LME to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

LME of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of LME. Such compounds may include agonists, antagonists, or partial or

inverse agonists. In one embodiment, an assay is performed under conditions permissive for LME activity, wherein LME is combined with at least one test compound, and the activity of LME in the presence of a test compound is compared with the activity of LME in the absence of the test compound. A change in the activity of LME in the presence of the test compound is indicative of a compound that modulates the activity of LME. Alternatively, a test compound is combined with an in vitro or cell-free system comprising LME under conditions suitable for LME activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of LME may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding LME or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding LME may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding LME can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding LME is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and

treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress LME, e.g., by secreting LME in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of LME and lipid metabolism enzymes. In addition, the expression of LME is closely associated with diseased, cancerous, inflamed, cardiovascular, and neurological tissues. Therefore, LME appears to play a role in cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders. In the treatment of disorders associated with increased LME expression or activity, it is desirable to decrease the expression or activity of LME. In the treatment of disorders associated with decreased LME expression or activity, it is desirable to increase the expression or activity of LME.

Therefore, in one embodiment, LME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LME. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia,

Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; and a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse,

rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation.

In another embodiment, a vector capable of expressing LME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LME including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified LME in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LME including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of LME may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LME including, but not limited to, those listed above.

In a further embodiment, an antagonist of LME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LME. Examples of such disorders include, but are not limited to, those cancers, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders described above. In one aspect, an antibody which specifically binds LME may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express LME.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding LME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LME including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of LME may be produced using methods which are generally known in the art. In particular, purified LME may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind LME. Antibodies to LME may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with LME or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to LME have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of LME amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to LME may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate

antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce LME-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for LME may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between LME and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering LME epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for LME. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of LME-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple LME epitopes, represents the average affinity, or avidity, of the antibodies for LME. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular LME epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the LME-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar

procedures which ultimately require dissociation of LME, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of LME-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding LME, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding LME. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding LME. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding LME may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

(Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in LME expression or regulation causes disease, the expression of LME from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in LME are treated by constructing mammalian expression vectors encoding LME and introducing these vectors by mechanical means into LME-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of LME include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). LME may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding LME from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to LME expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding LME under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding LME to cells which have one or more genetic abnormalities with respect to the expression of LME. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"),

hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding LME to target cells which have one or more genetic abnormalities with respect to the expression of LME. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing LME to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding LME to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for LME into the alphavirus genome in place of the capsid-coding region results in the production of a large number of LME-coding RNAs and the synthesis of high levels of LME in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application

(Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of LME into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding LME.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding LME. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding LME. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased LME expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding LME may be therapeutically useful, and in the treament of disorders associated with decreased LME expression or activity, a compound which specifically promotes expression of the polynucleotide encoding LME may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding LME is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding LME are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding LME. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the

polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of LME, antibodies to LME, and mimetics, agonists, antagonists, or inhibitors of LME.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration

without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising LME or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, LME or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example LME or fragments thereof, antibodies of LME, and agonists, antagonists or inhibitors of LME, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind LME may be used for the diagnosis of disorders characterized by expression of LME, or in assays to monitor patients being treated with LME or agonists, antagonists, or inhibitors of LME. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for LME include methods which utilize the antibody and a label to detect LME in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring LME, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of LME expression. Normal or standard values for LME expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to LME under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of LME expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding LME may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of LME may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of LME, and to monitor regulation of LME levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding LME or closely related molecules may be used to identify nucleic acid sequences which encode LME. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding LME, allelic variants, or related

sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the LME encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:6-10 or from genomic sequences including promoters, enhancers, and introns of the LME gene.

Means for producing specific hybridization probes for DNAs encoding LME include the cloning of polynucleotide sequences encoding LME or LME derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding LME may be used for the diagnosis of disorders associated with expression of LME. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal

degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; and a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective

endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation . The polynucleotide sequences encoding LME may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered LME expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding LME may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding LME may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding LME in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of LME, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding LME, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from

samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding LME may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding LME, or a fragment of a polynucleotide complementary to the polynucleotide encoding LME, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding LME may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding LME are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation

of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of LME include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, LME, fragments of LME, or antibodies specific for LME may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The

resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles,

are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for LME to quantify the levels of LME expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated

biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding LME may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic

map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding LME on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, LME, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between LME and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with LME, or fragments thereof, and washed. Bound LME is then detected by methods well known in the art. Purified LME can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding LME specifically compete with a test compound for binding LME. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with LME.

In additional embodiments, the nucleotide sequences which encode LME may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

Table 6

Library	Vector	Library Description
ADRETUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from left upper pole, adrenal gland tumor tissue removed from a 52-year-old Caucasian male during nephroureterectomy and local destruction of renal lesion. Pathology indicated grade 3 adrenal cortical carcinoma forming a mass that infiltrated almost the whole adrenal parenchyma and extended to adjacent adipose tissue. A metastatic tumor nodule was identified in the hilar region. The renal vein was infiltrated by tumor and the neoplastic process was present at the resection margin of the renal vein. Fragments of adrenal cortical carcinoma and thrombus were found in the inferior vena cava. Patient history included abnormal weight loss. Family history included skin cancer, type I diabetes, and neurotic depression.
BLADNOT04	pINCY	Library was constructed using RNA isolated from bladder tissue of a 28-year-old Caucasian male, who died from a self-inflicted gunshot wound..
COLRTUE01	PSPORT1	This 5' biased random primed library was constructed using RNA isolated from rectum tumor tissue removed from a 50-year-old Caucasian male during closed biopsy of rectum and resection of rectum. Pathology indicated grade 3 colonic adenocarcinoma which invades through the muscularis propria to involve pericolonic fat. Tubular adenoma with low grade dysplasia was also identified. The patient presented with malignant rectal neoplasm, blood in stool, and constipation. Patient history included benign neoplasm of the large bowel, hyperlipidemia, benign hypertension, alcohol abuse, and tobacco abuse. Previous surgeries included above knee amputation and vasectomy. Patient medications included allopurinol, Zantac, Darvocet, Centrum vitamins, and an unspecified stool softener. Family history included congestive heart failure in the mother; and benign neoplasm of the large bowel and polypectomy in the sibling(s).

Table 6 (cont.)

Library	Vector	Library Description
PITUNON01	pINCY	This normalized pituitary gland tissue library was constructed from 6.92 million independent clones from a pituitary gland tissue library. Starting RNA was made from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5,

b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-5.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:6-10.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10,

b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b),

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

18. A method for treating a disease or condition associated with decreased expression of functional LME, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional LME, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional LME, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of LME in a

properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/203,511, U.S. Ser. No. 60/207,903, U.S. Ser. No. 60/210,150, and U.S. Ser. No. 60/213,392, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid

(Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and

programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:6-10. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative lipid metabolism enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode lipid metabolism enzymes, the encoded polypeptides were analyzed by querying against PFAM models for lipid metabolism enzymes. Potential lipid metabolism enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as lipid metabolism enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated

but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of LME Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:6-10 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:6-10 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between

chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding LME are analyzed with respect to the

tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding LME. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of LME Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:6-10 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06

software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is

described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are

differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the LME-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring LME. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of LME. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the LME-encoding transcript.

XII. Expression of LME

Expression and purification of LME is achieved using bacterial or virus-based expression systems. For expression of LME in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express LME upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of LME in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding LME by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong

polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, LME is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from LME at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified LME obtained by these methods can be used directly in the assays shown in Examples XVI and XVII, where applicable.

XIII. Functional Assays

LME function is assessed by expressing the sequences encoding LME at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake;

alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of LME on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding LME and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding LME and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of LME Specific Antibodies

LME substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the LME amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-LME activity by, for example, binding the peptide or LME to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring LME Using Specific Antibodies

Naturally occurring or recombinant LME is substantially purified by immunoaffinity chromatography using antibodies specific for LME. An immunoaffinity column is constructed by covalently coupling anti-LME antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing LME are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of LME (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/LME binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and LME is collected.

XVI. Identification of Molecules Which Interact with LME

LME, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled LME, washed, and any wells with labeled LME complex are assayed. Data obtained using different concentrations of LME are used to calculate values for the number, affinity, and association of LME with the candidate molecules.

Alternatively, molecules interacting with LME are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

LME may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of LME Activity

LME activity can be demonstrated by an in vitro hydrolysis assay with vesicles containing 1-palmitoyl-2-[1-^{14}C]oleoyl phosphatidylcholine (Sigma-Aldrich). LME triglyceride lipase activity and phospholipase A_2 activity are demonstrated by analysis of the cleavage products isolated from the hydrolysis reaction mixture.

Vesicles containing 1-palmitoyl-2-[1-^{14}C]oleoyl phosphatidylcholine (Amersham Pharmacia Biotech.) are prepared by mixing 2.0 µCi of the radiolabeled phospholipid with 12.5 mg of unlabeled 1-palmitoyl-2-oleoyl phosphatidylcholine and drying the mixture under N_2. 2.5 ml of 150 mM Tris-HCl, pH 7.5, is added, and the mixture is sonicated and centrifuged. The supernatant may be stored at 4 °C. The final reaction mixtures contain 0.25 ml of Hanks buffered salt solution supplemented with 2.0 mM taurochenodeoxycholate, 1.0% bovine serum albumin, 1.0 mM $CaCl_2$, pH 7.4, 150 µg of 1-palmitoyl-2-[1-^{14}C]oleoyl phosphatidylcholine vesicles, and various amounts of LME diluted in PBS. After incubation for 30 min at 37 °C, 20 µg each of lyso-phosphatidylcholine and oleic acid are added as carriers and each sample is extracted for total lipids. The lipids are separated by thin layer chromatography using a two solvent system of chloroform:methanol:acetic acid:water (65:35:8:4) until the solvent front is halfway up the plate. The process is then continued

with hexane:ether:acetic acid (86:16:1) until the solvent front is at the top of the plate. The lipid-containing areas are visualized with I_2 vapor; the spots are scraped, and their radioactivity is determined by scintillation counting. The amount of radioactivity released as fatty acids will increase as a greater amount of LME is added to the assay mixture while the amount of radioactivity released as lyso-phosphatidylcholine will remain low. This demonstrates that LME cleaves at the sn-2 and not the sn-1 position, as is characteristic of phospholipase A_2 activity.

Alternatively, LME phospholipase activity is measured by the hydrolysis of a fatty acyl residue at the sn-1 position of phosphatidylserine. LME is combined with the Tritium [^{3}H] labeled substrate phosphatidylserine at stoichiometric quantities in a suitable buffer. Following an appropriate incubation time, the hydrolyzed reaction products are separated from the substrates by chromatographic methods. The amount of acylglycerophosphoserine produced is measured by counting tritiated product with the help of a scintillation counter. Various control groups are set up to account for background noise and unincorporated substrate. The final counts represent the tritiated enzyme product [^{3}H]-acylglycerophosphoserine, which is directly proportional to the activity of LME in biological samples.

LME lipoxygenase activity can be measured by chromatographic methods. Extracted LME lipoxygenase protein (200 μg) is incubated with 100 μM [1-^{14}C] arachidonic acid or other unlabeled fatty acids at 37°C for 15-30 min. After the incubation, stop solution (acetonitrile:methanol:water, 350:150:1) is added. The samples are extracted and analyzed by reverse-phase HPLC by using a solvent system of methanol/water/acetic acid, 85:15:0.01 (vol/vol) at a flow rate of 1 ml/min. The effluent is monitored at 235 nm and analyzed for the presence of the major arachidonic metabolite such as 12-HPETE (catalyzed by 12-LOX). The fractions are also subjected to liquid scintillation counting. The final counts represent the products, which is directly proportional to the activity of LME in biological samples. For stereochemical analysis, the metabolites of arachidonic acid are analyzed further by chiral phase-HPLC and by mass spectrometry (Sun, D. et al. (1998) J. Biol. Chem. 273:33540-33547).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7473568	1	7473568CD1	6	7473568CB1
7473224	2	7473224CD1	7	7473224CB1
7473234	3	7473234CD1	8	7473234CB1
1321517	4	1321517CD1	9	1321517CB1
7472768	5	7472768CD1	10	7472768CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	7473568CD1	g304241	0	[Bos taurus] phospholipase C (Ferreira,P.A., et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90: 6042-6046)
2	7473224CD1	g6174881	6.9e-74	[Mus musculus] group IIF secreted phospholipase A2 (Valentin,E., et al. (1999) J. Biol. Chem. 274: 31195-31202)
		g12276060	1.00E-104	[fl][Homo sapiens] group IIF secreted phospholipase A2 (Valentin,-E., et al. (2000) Biochem. Biophys. Res. Commun. 279: 223-228)
3	7473234CD1	g10441004	0	[fl][Homo sapiens] epidermal lipoxygenase
		g5304928	0	lipoxygenase-3 [Mus musculus] (Kinzig,A. et al. (1999) Genomics 58: 158-164)
4	1321517CD1	g163522	0	[Bos taurus] phospholipase C (Katan M. et al. (1988) Cell 54:171-177)
5	7472768CD1	g6705987	2.1e-150	[Mus musculus] phospholipase C-L2 (Otsuki,M. et al. (1999) Biochem. Biophys. Res. Commun. 266: 97-103)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7473568CD1	1093	S62 T65 S80 T100 T220 S421 S450 S613 T632 S699 T744 T768 T779 S840 S995 T1014 S1086 S426 T897 T107 S228 S337 S366 S408 T440 S493 T592 T619 S742 S769 S788 S802 T858 S872 S931 T1063 Y123	N490 N495 N567 N1010 N1084	Phosphatidylinositol-specific phospholipase PI-PLC-X:E201-R351, PI-PLC-Y:L464-R580	HMMER_PFAM
					C2 domain: C602-I685	
					Phosphatidylinositol-specific phospholipase X-box domain BL50007A:L206-G251 BL50007B:C267-Q304 BL50007C:L335-R351 BL50007D:Y513-D554 BL50007E:A672-L708	BLIMPS_BLOCKS
					Phospholipase C signature PR00390A:P205-Q223 PR00390B:E231-G251 PR00390C:A334-R351 PR00390D:M518-W539 PR00390E:W539-M557 PR00390F:L686-R696	BLIMPS_PRINTS
					C2 domain signature PR00360A:S613-M625 PR00360B:N645-F658 PR00360C:L669-N677	
					1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D DM00855|A48047|58-521:M1-S408 DM00712|A48047|523-820:A409-L719 DM00855|P13217|63-512:M1-A397 DM00712|A53766|83-369:Y453-L719	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					BETA PHOSPHOLIPASE 1 PHOSPHATIDYL-INOSITOL4 5 BISPHOSPHATE PHOSPHODIESTERASE PD023749:E355-Y463	BLAST_PRODOM
					PHOSPHOLIPASE C PHOSPHODIESTERASE 1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PD001214:E201-R351	
					PHOSPHOLIPASE 1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PHOSPHODIESTERASE PD011437:L952-K1048	
					PHOSPHOLIPASE PHOSPHODIESTERASE C HYDROLASE 1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PHOSPHOINOSITIDESPECIFIC PD001202:L464-R580	
					Phospholipase A2 histidine active site: C106-C113	MOTIFS
2	7473224CD1	211	T147 T168 S189 T37 T76 T168	N135 N145 N166 N187	signal_cleavage:M1-G63	SPSCAN
					Phospholipase A2: S64-C188	HMMER_PFAM
					Phospholipase A2 histidine proteins BL00118A:S64-T76 BL00118B:Y87-Y114 BL00118C:C122-V140 BL00118D:C149-L164	BLIMPS_BLOCKS
					Phospholipase A2 active sites signatures pa2_his.prf:G86-N135	PROFILESCAN
					PHOSPHOLIPASE A2 SIGNATURE PR00389B:A80-Q98 PR00389C:P99-L117 PR00389E:C149-M165	BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					A2 PHOSPHOLIPASE HYDROLASE PHOSPHATIDYLCHOLINE LIPID DEGRADATION CALCIUM VENOM MULTIGENE FAMILY PD000303:S64-Q182	BLAST_PRODOM
					PHOSPHOLIPASE A2 ASPARTIC ACID DM00093|P48076|21-138:S64-R174 DM00093|P04417|1-112:L65-V179 DM00093|P39877|20-134:G63-V179 DM00093|P49121|17-128:S64-F176	BLAST_DOMO
3	7473234	711	S161 S167 S242 S33 S384 S494 S608 S81 T117 T178 T260 T29 T316 T382 T423 T569 T621 T624 T628 T657 T698 T90	N21 N405 N583 N633	Lipoxygenase: H267-P703	HMMER_PFAM
					PLAT/LH2 domain: A2-G116	
					Lipoxygenases iron-binding region proteins BL00711A:F99-Y108 BL00711C:W268-K296 BL00711D:A347-P372 BL00711E:L414-I450 BL00711F:F484-V500 BL00711G:I503-D534 BL00711H:A535-M573 BL00711I:L577-P614	BLIMPS_BLOCKS
					Lipoxygenase signature PR00087A:E385-V402 PR00087B:H403-A420 PR00087C:T423-Y443	BLIMPS_PRINTS
					Mammalian lipoxygenase signature PR00467A:G11-G28 PR00467B:L61-D80 PR00467C:R138-G152 PR00467D:W243-V264 PR00467E:Q344-P363 PR00467G:T628-V645	

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					OXIDOREDUCTASE DIOXYGENASE IRON LIPOXYGENASE ARACHIDONATE LEUKOTRIENE BIOSYNTHESIS MULTIGENE FAMILY 12 LIPOXYGENASE PD000872:G373-L695	BLAST_PRODOM
					ARACHIDONATE 12 LIPOXYGENASE EPIDERMIS TYPE 12R LIPOXYGENASE 15S LIPOXYGENASE 8S LIPOXYGENASE 8R LIPOXYGENASEALLENE OXIDE PD150360:M1-T178	
					LIPOXYGENASES IRON-BINDING REGION DM00593|P09917|329-672:Q370-I711 DM00593|P18054|319-661:P372-I711 DM00593|P16050|319-660:L368-I711 DM00593|B54075|321-663:L368-I711	BLAST_DOMO
					Lipoxygenases iron-binding region signatures lipoxygenase_2.prf:L410-V462	PROFILESCAN
4	1321517	1216	S196 S197 S21 S300 S344 S387 S417 S454 S470 S477 S483 S495 S511 S523 S524 S558 S569 S582 S679 S689 S887 S969 S974 S978 S981 S988 S1026 S1089 S1118 S1199 T26 T369 T489 T509 T602 T695 T708 T802 T831 T847 T858 T871 T903 T909 T942 T949 T1002 Y263 Y589	N113 N145 N199 N281 N282 N328 N343 N562 N865	Phosphatidylinositol-specific phospholipase C PI-PLC-X:D317-K468 PI-PLC-Y:E539-R656	HMMER_PFAM
					C2 domain:L678-I761	
					Phosphatidylinositol-specific phospholipase X-box domain BL50007A:L322-G367 BL50007B:T383-Q420 BL50007C:L452-K468 BL50007D:Y589-D630 BL50007E:A748-L784	BLIMPS_BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PHOSPHOLIPASE C SIGNATURE PR00390A:P321-Q339 PR00390B:E347-G367 PR00390C:P451-K468 PR00390D:L594-W615 PR00390E:W615-M633 PR00390F:L762-H772	BLIMPS_PRINTS
					C2 domain signature PR00360A:S689-M701 PR00360B:N721-F734 PR00360C:L745-G753	
					1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PHOSPHODIESTERASE BETA PLC PD155902:A970-L1216	BLAST_PRODOM
					PHOSPHOLIPASE C PHOSPHODIESTERASE HYDROLASE 1 PHOSPHATIDYLINOSITOL 4 5BISPHOSPHATE LIPID DEGRADATION PHOSPHOINOSITIDE-SPECIFIC PD001214:D317-K472	
					PHOSPHOLIPASE HYDROLASE BETA 1 PHOSPHATIDYLINOSITOL4 5 BISPHOSPHATE PHOSPHODIESTERASE C LIPID DEGRADATION PD006352:V18-K172	

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PHOSPHOLIPASE BETA HYDROLASE 1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PHOSPHODIESTERASE LIPID DEGRADATION PD004886:R760-H866	BLAST_PRODOM
					1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D DM00855|P10894|62-503:E62-A504 DM00855|A48001|62-505:E62-S501 DM00855|Q01970|62-505:E62-P502 DM00712|P10894|505-794:G505-V795	BLAST_DOMO
5	7472768	1239	S113 S185 S267 S308 S38 S382 S477 S50 S515 S519 S575 S582 S590 S597 S609 S616 S623 S681 S713 S83 S913 S950 S964 T172 T268 T419 T536 T544 T647 T844 T93 Y660 S1132 S1012 S1143 S1212	N322 N335 N504 N566	Phosphatidylinositol-specific phospholipase PI-PLC-X:D355-K500 PI-PLC-Y:A653-C768	HMMER_PFAM
					C2 domain:L788-T880	
					PH domain:A3-A110	
					Phosphatidylinositol-specific phospholipase X-box domain proteins BL50007A:L360-G405 BL50007B:T419-Q456 BL50007C:L484-K500 BL50007D:F701-G742 BL50007E:D867-I903	BLIMPS_BLOCKS
					Phospholipase C signature PR00390A:P359-Q377 PR00390B:D385-G405 PR00390C:T483-K500 PR00390D:L706-W727 PR00390E:W727-L745 PR00390F:L881-R891	BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					C2 domain signature PR00360A:R809-I821 PR00360B:N839-M852 PR00360C:V861-D869	BLIMPS_PRINTS
					PHOSPHOLIPASE C 1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PHOSPHOINOSITIDE SPECIFIC PD001214:D355-K500	BLAST_PRODOM
					PHOSPHOLIPASE C 1PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE PHOSPHOINOSITIDE SPECIFIC PD001202:L654-P764	
					PHOSPHOLIPASE 1 PHOSPHATIDYLINOSITOL 4 5 BISPHOSPHATE CALCIUM BINDING PD004439:Q243-Q354	
					1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D DM00855|P51178|64-472: D249-D526, S47-R123, P142-A215 DM00855|P08487|71-500: L242-G532, I46-M109, G150-R204 DM00855|P16885|63-486: H250-E517, I46-L108, W168-H214 DM00855|A53970|67-522: Y279-P502, I46-M109	BLAST_DOMO
					Ef_Hand: D177-V189	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
6	7473568CB1	4813	3490-3552, 1788-1819, 1537-1679, 1-1030, 4728-4813, 2237-2540	7412054H1 (BONMTUE02)	334	915
				2496596H1 (ADRETUT05)	3431	3669
				70559575V1	1133	1672
				3715092H1 (PENCNOT09)	4519	4813
				2763231F6 (BRSTNOT12)	1813	2343
				70560443V1	969	1511
				6200132H1 (PITUNON01)	3611	4277
				70559016V1	1628	2292
				2408864H1 (BSTMNON02)	4371	4608
				484909R6 (HNT2RAT01)	641	1121
				70559065V1	2990	3643
				7606844H1 (COLRTUE01)	1	652
				70559164V1	2116	2899
				70559183V1	2459	3113
				3719593H1 (PENCNOT10)	4441	4714
				5768362H1 (STOMFET02)	3755	4390
7	7473224CB1	2270	1-90, 482-2270	71227816V1	1	535
				71413435V1	533	1233
				70862340V1	1306	1870
				70863240V1	1215	1795
				70862953V1	1681	2270
				70864195V1	416	1038
8	7473234CB1	2136	1-1026,	FL1078033_00001	1	2136
9	1321517CB1	3700	3212-3290, 647-1124, 1-55, 3384-3700, 1450-1988, 2345-2791	CpG_NWW300397336.R1	1	314
				g4958835	914	1819
				g3947829	3426	3648
				g5514658	299	940
				g6807699	831	3475
				1321517H1 (BLADNOT04)	846	1035
				71200138V1	1	618
				71200194V1	44	675
				GBI.g3947829.edit1	3648	3700

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
10	7472768CB1	4237	1-413, 730-826, 2386-2573, 480-597, 3136-4237, 1841-1899"	GBI:g8039388.edit1	1	135
				7314090H1 (LIVRNOE07)	135	480
				GBI:g8039388.edit1	480	1020
				7690378J1 (PROSTME06)	3650	4237
				7647838H1 (UTRSTUE01)	3280	3910
				7673955H2 (NOSETUE01)	2129	2659
				7089668H1 (BRAUTDR03)	3037	3533
				8000203H1 (LNODTUC02)	2784	3141
				7617171H1 (KIDNTUE01)	2535	3077
				6199592F8. (PITUNON01)	1020	1592
				55047574H1	1411	2034
				GNN:g6983515_000028_002	2034	2129

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
6	7473568CB1	COLRTUE01
7	7473224CB1	PROSTUT09
8	7473234CB1	ADRETUR01
9	1321517CB1	BLADNOT04
10	7472768CB1	PITUNON01

biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of LME in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of LME in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:6.

51. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:7.

52. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:8.

53. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:9.

54. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:10.

<110> INCYTE GENOMICS, INC.
DAS, Debopriya
REDDY, Roopa
YAO, Monique G.
NGUYEN, Danniel B.
LU, Yan
TRIBOULEY, Catherine M.
YUE, Henry
KHAN, Farrah A.
GANDHI, Ameena R.
AU-YOUNG, Janice
LAL, Preeti
KEARNEY, Liam
ELLIOTT, Vicki S.
DING, Li
THORNTON, Michael

<120> LIPID METABOLISM ENZYMES

<130> PI-0095 PCT

<140> To Be Assigned
<141> Herewith

<150> 60,203,511; 60/207,903; 60/210,150; 60/213,392
<151> 2000-05-11; 2000-05-25; 2000-06-07; 2000-06-23

<160> 10
<170> PERL Program

<210> 1
<211> 1093
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473568CD1

<400> 1
Met Val Ala Glu Asn Pro Glu Val Thr Lys Gln Trp Val Glu Gly
1 5 10 15
Leu Arg Ser Ile Ile His Asn Phe Arg Ala Asn Asn Val Ser Pro
20 25 30
Met Thr Cys Leu Lys Lys His Trp Met Lys Leu Ala Phe Met Thr
35 40 45
Asn Thr Asn Gly Lys Ile Pro Val Arg Ser Ile Thr Arg Thr Phe
50 55 60
Ala Ser Gly Lys Thr Glu Lys Val Ile Phe Gln Ala Leu Lys Glu
65 70 75
Leu Gly Leu Pro Ser Gly Lys Asn Asp Glu Ile Glu Pro Thr Ala
80 85 90
Phe Ser Tyr Glu Lys Phe Tyr Glu Leu Thr Gln Lys Ile Cys Pro
95 100 105
Arg Thr Asp Ile Glu Asp Leu Phe Lys Lys Ile Asn Gly Asp Lys
110 115 120
Thr Asp Tyr Leu Thr Val Asp Gln Leu Val Ser Phe Leu Asn Glu
125 130 135
His Gln Arg Asp Pro Arg Leu Asn Glu Ile Leu Phe Pro Phe Tyr
140 145 150
Asp Ala Lys Arg Ala Met Gln Ile Ile Glu Met Tyr Glu Pro Asp
155 160 165

Glu Asp Leu Lys Lys Lys Gly Leu Ile Ser Ser Asp Gly Phe Cys
170 175 180
Arg Tyr Leu Met Ser Asp Glu Asn Ala Pro Val Phe Leu Asp Arg
185 190 195
Leu Glu Leu Tyr Gln Glu Met Asp His Pro Leu Ala His Tyr Phe
200 205 210
Ile Ser Ser Ser His Asn Thr Tyr Leu Thr Gly Arg Gln Phe Gly
215 220 225
Gly Lys Ser Ser Val Glu Met Tyr Arg Gln Val Leu Leu Ala Gly
230 235 240
Cys Arg Cys Val Glu Leu Asp Cys Trp Asp Gly Lys Gly Glu Asp
245 250 255
Gln Glu Pro Ile Ile Thr His Gly Lys Ala Met Cys Thr Asp Ile
260 265 270
Leu Phe Lys Asp Val Ile Gln Ala Ile Lys Glu Thr Ala Phe Val
275 280 285
Thr Ser Glu Tyr Pro Val Ile Leu Ser Phe Glu Asn His Cys Ser
290 295 300
Lys Tyr Gln Gln Tyr Lys Met Ser Lys Tyr Cys Glu Asp Leu Phe
305 310 315
Gly Asp Leu Leu Leu Lys Gln Ala Leu Glu Ser His Pro Leu Glu
320 325 330
Pro Gly Arg Ala Leu Pro Ser Pro Asn Asp Leu Lys Arg Lys Ile
335 340 345
Leu Ile Lys Asn Lys Arg Leu Lys Pro Glu Val Glu Lys Lys Gln
350 355 360
Leu Glu Ala Leu Arg Ser Met Met Glu Ala Gly Glu Ser Ala Ser
365 370 375
Pro Ala Asn Ile Leu Glu Asp Asp Asn Glu Glu Glu Ile Glu Ser
380 385 390
Ala Asp Gln Glu Glu Glu Ala His Pro Glu Phe Lys Phe Gly Asn
395 400 405
Glu Leu Ser Ala Asp Asp Leu Gly His Lys Glu Ala Val Ala Asn
410 415 420
Ser Val Lys Lys Ala Ser Asp Asp Leu Glu His Glu Asn Asn Lys
425 430 435
Lys Gly Leu Val Thr Val Glu Asp Glu Gln Ala Trp Met Ala Ser
440 445 450
Tyr Lys Tyr Val Gly Ala Thr Thr Asn Ile His Pro Tyr Leu Ser
455 460 465
Thr Met Ile Asn Tyr Ala Gln Pro Val Lys Phe Gln Gly Phe His
470 475 480
Val Ala Glu Glu Arg Asn Ile His Tyr Asn Met Ser Ser Phe Asn
485 490 495
Glu Ser Val Gly Leu Gly Tyr Leu Lys Thr His Ala Ile Glu Phe
500 505 510
Val Asn Tyr Asn Lys Arg Gln Met Ser Arg Ile Tyr Pro Lys Gly
515 520 525
Gly Arg Val Asp Ser Ser Asn Tyr Met Pro Gln Ile Phe Trp Asn
530 535 540
Ala Gly Cys Gln Met Val Ser Leu Asn Tyr Gln Thr Pro Asp Leu
545 550 555
Ala Met Gln Leu Asn Gln Gly Lys Phe Glu Tyr Asn Gly Ser Cys
560 565 570
Gly Tyr Leu Leu Lys Pro Asp Phe Met Arg Arg Pro Asp Arg Thr
575 580 585
Phe Asp Pro Phe Ser Glu Thr Pro Val Asp Gly Val Ile Ala Ala
590 595 600
Thr Cys Ser Val Gln Val Ile Ser Gly Gln Phe Leu Ser Asp Lys
605 610 615
Lys Ile Gly Thr Tyr Val Glu Val Asp Met Tyr Gly Leu Pro Thr
620 625 630
Asp Thr Ile Arg Lys Glu Phe Arg Thr Arg Met Val Met Asn Asn

635 640 645
Gly Leu Asn Pro Val Tyr Asn Glu Glu Ser Phe Val Phe Arg Lys
650 655 660
Val Ile Leu Pro Asp Leu Ala Val Leu Arg Ile Ala Val Tyr Asp
665 670 675
Asp Asn Asn Lys Leu Ile Gly Gln Arg Ile Leu Pro Leu Asp Gly
680 685 690
Leu Gln Ala Gly Tyr Arg His Ile Ser Leu Arg Asn Glu Gly Asn
695 700 705
Lys Pro Leu Ser Leu Pro Thr Ile Phe Cys Asn Ile Val Leu Lys
710 715 720
Thr Tyr Val Pro Asp Gly Phe Gly Asp Ile Val Asp Ala Leu Ser
725 730 735
Asp Pro Lys Lys Phe Leu Ser Ile Thr Glu Lys Arg Ala Asp Gln
740 745 750
Met Arg Ala Met Gly Ile Glu Thr Ser Asp Ile Ala Asp Val Pro
755 760 765
Ser Asp Thr Ser Lys Asn Asp Lys Lys Gly Lys Ala Asn Thr Ala
770 775 780
Lys Ala Asn Val Thr Pro Gln Ser Ser Ser Glu Leu Arg Pro Thr
785 790 795
Thr Thr Ala Ala Leu Ala Ser Gly Val Glu Ala Lys Lys Gly Ile
800 805 810
Glu Leu Ile Pro Gln Val Arg Ile Glu Asp Leu Lys Gln Met Lys
815 820 825
Ala Tyr Leu Lys His Leu Lys Lys Gln Gln Lys Glu Leu Asn Ser
830 835 840
Leu Lys Lys Lys His Ala Lys Glu His Ser Thr Met Gln Lys Leu
845 850 855
His Cys Thr Gln Val Asp Lys Ile Val Ala Gln Tyr Asp Lys Glu
860 865 870
Lys Ser Thr His Glu Lys Ile Leu Glu Lys Ala Met Lys Lys Lys
875 880 885
Gly Gly Ser Asn Cys Leu Glu Met Lys Lys Glu Thr Glu Ile Lys
890 895 900
Ile Gln Thr Leu Thr Ser Asp His Lys Ser Lys Val Lys Glu Ile
905 910 915
Val Ala Gln His Thr Lys Glu Trp Ser Glu Met Ile Asn Thr His
920 925 930
Ser Ala Glu Glu Gln Glu Ile Arg Asp Leu His Leu Ser Gln Gln
935 940 945
Cys Glu Leu Leu Lys Lys Leu Leu Ile Asn Ala His Glu Gln Gln
950 955 960
Thr Gln Gln Leu Lys Leu Ser His Asp Arg Glu Ser Lys Glu Met
965 970 975
Arg Ala His Gln Ala Lys Ile Ser Met Glu Asn Ser Lys Ala Ile
980 985 990
Ser Gln Asp Lys Ser Ile Lys Asn Lys Ala Glu Arg Glu Arg Arg
995 1000 1005
Val Arg Glu Leu Asn Ser Ser Asn Thr Lys Lys Phe Leu Glu Glu
1010 1015 1020
Arg Lys Arg Leu Ala Met Lys Gln Ser Lys Glu Met Asp Gln Leu
1025 1030 1035
Lys Lys Val Gln Leu Glu His Leu Glu Phe Leu Glu Lys Gln Asn
1040 1045 1050
Glu Gln Leu Leu Lys Ser Cys His Ala Val Ser Gln Thr Gln Gly
1055 1060 1065
Glu Gly Asp Ala Ala Asp Gly Glu Ile Gly Ser Arg Asp Gly Pro
1070 1075 1080
Gln Thr Ser Asn Ser Ser Met Lys Leu Gln Asn Ala Asn
1085 1090

<210> 2

<211> 211
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473224CD1

<400> 2
Met Ala Asp Gly Ala Lys Ala Asn Pro Lys Gly Phe Lys Lys Lys
1 5 10 15
Val Leu Asp Arg Cys Phe Ser Gly Trp Arg Gly Pro Arg Phe Gly
20 25 30
Ala Ser Cys Pro Ser Arg Thr Ser Arg Ser Ser Leu Gly Met Lys
35 40 45
Lys Phe Phe Thr Val Ala Ile Leu Ala Gly Ser Val Leu Ser Thr
50 55 60
Ala His Gly Ser Leu Leu Asn Leu Lys Ala Met Val Glu Ala Val
65 70 75
Thr Gly Arg Ser Ala Ile Leu Ser Phe Val Gly Tyr Gly Cys Tyr
80 85 90
Cys Gly Leu Gly Gly Arg Gly Gln Pro Lys Asp Glu Val Asp Trp
95 100 105
Cys Cys His Ala His Asp Cys Cys Tyr Gln Glu Leu Phe Asp Gln
110 115 120
Gly Cys His Pro Tyr Val Asp His Tyr Asp His Thr Ile Glu Asn
125 130 135
Asn Thr Glu Ile Val Cys Ser Asp Leu Asn Lys Thr Glu Cys Asp
140 145 150
Lys Gln Thr Cys Met Cys Asp Lys Asn Met Val Leu Cys Leu Met
155 160 165
Asn Gln Thr Tyr Arg Glu Glu Tyr Arg Gly Phe Leu Asn Val Tyr
170 175 180
Cys Gln Gly Pro Thr Pro Asn Cys Ser Ile Tyr Glu Pro Pro Pro
185 190 195
Glu Glu Val Thr Cys Ser His Gln Ser Pro Ala Pro Pro Ala Pro
200 205 210
Pro

<210> 3
<211> 711
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473234CD1

<400> 3
Met Ala Val Tyr Arg Leu Cys Val Thr Thr Gly Pro Tyr Leu Arg
1 5 10 15
Ala Gly Thr Leu Asp Asn Ile Ser Val Thr Leu Val Gly Thr Cys
20 25 30
Gly Glu Ser Pro Lys Gln Arg Leu Asp Arg Met Gly Arg Asp Phe
35 40 45
Ala Pro Gly Ser Val Gln Lys Tyr Lys Val Arg Cys Thr Ala Glu
50 55 60
Leu Gly Glu Leu Leu Leu Leu Arg Val His Lys Glu Arg Tyr Ala
65 70 75
Phe Phe Arg Lys Asp Ser Trp Tyr Cys Ser Arg Ile Cys Val Thr
80 85 90
Glu Pro Asp Gly Ser Val Ser His Phe Pro Cys Tyr Gln Trp Ile

```
                95                  100                 105
Glu Gly Tyr Cys Thr Val Glu Leu Arg Pro Gly Thr Ala Arg Thr
            110                 115                 120
Ile Cys Gln Asp Ser Leu Pro Leu Leu Leu Asp His Arg Thr Arg
        125                 130                 135
Glu Leu Arg Ala Arg Gln Glu Cys Tyr Arg Trp Lys Ile Tyr Ala
    140                 145                 150
Pro Gly Phe Pro Cys Met Val Asp Val Asn Ser Phe Gln Glu Met
155                 160                 165
Glu Ser Asp Lys Lys Phe Ala Leu Thr Lys Thr Thr Thr Cys Val
            170                 175                 180
Asp Gln Gly Asp Ser Ser Gly Asn Arg Tyr Leu Pro Gly Phe Pro
        185                 190                 195
Met Lys Ile Asp Ile Pro Ser Leu Met Tyr Met Glu Pro Asn Val
    200                 205                 210
Arg Tyr Ser Ala Thr Lys Thr Ile Ser Leu Leu Phe Asn Ala Ile
215                 220                 225
Pro Ala Ser Leu Gly Met Lys Leu Arg Gly Leu Leu Asp Arg Lys
            230                 235                 240
Gly Ser Trp Lys Lys Leu Asp Asp Met Gln Asn Ile Phe Trp Cys
        245                 250                 255
His Lys Thr Phe Thr Thr Lys Tyr Val Thr Glu His Trp Cys Glu
    260                 265                 270
Asp His Phe Phe Gly Tyr Gln Tyr Leu Asn Gly Val Asn Pro Val
275                 280                 285
Met Leu His Cys Ile Ser Ser Leu Pro Ser Lys Leu Pro Val Thr
            290                 295                 300
Asn Asp Met Val Ala Pro Leu Leu Gly Gln Asp Thr Cys Leu Gln
305                 310                 315
Thr Glu Leu Glu Arg Gly Asn Ile Phe Leu Ala Asp Tyr Trp Ile
            320                 325                 330
Leu Ala Glu Ala Pro Thr His Cys Leu Asn Gly Arg Gln Gln Tyr
        335                 340                 345
Val Ala Ala Pro Leu Cys Leu Leu Trp Leu Ser Pro Gln Gly Ala
    350                 355                 360
Leu Val Pro Leu Ala Ile Gln Leu Ser Gln Thr Pro Gly Pro Asp
365                 370                 375
Ser Pro Ile Phe Leu Pro Thr Asp Ser Glu Trp Asp Trp Leu Leu
            380                 385                 390
Ala Lys Thr Trp Val Arg Asn Ser Glu Phe Leu Val His Glu Asn
        395                 400                 405
Asn Thr His Phe Leu Cys Thr His Leu Leu Cys Glu Ala Phe Ala
    410                 415                 420
Met Ala Thr Leu Arg Gln Leu Pro Leu Cys His Pro Ile Tyr Lys
425                 430                 435
Leu Leu Leu Pro His Thr Arg Tyr Thr Leu Gln Val Asn Thr Ile
            440                 445                 450
Ala Arg Ala Thr Leu Leu Asn Pro Glu Gly Leu Val Asp Gln Val
        455                 460                 465
Thr Ser Ile Gly Arg Gln Gly Leu Ile Tyr Leu Met Ser Thr Gly
    470                 475                 480
Leu Ala His Phe Thr Tyr Thr Asn Phe Cys Leu Pro Asp Ser Leu
485                 490                 495
Arg Ala Arg Gly Val Leu Ala Ile Pro Asn Tyr His Tyr Arg Asp
            500                 505                 510
Asp Gly Leu Lys Ile Trp Ala Ala Ile Glu Ser Phe Val Ser Glu
        515                 520                 525
Ile Val Gly Tyr Tyr Tyr Pro Ser Asp Ala Ser Val Gln Gln Asp
    530                 535                 540
Ser Glu Leu Gln Ala Trp Thr Gly Glu Ile Phe Ala Gln Ala Phe
545                 550                 555
Leu Gly Arg Glu Ser Ser Gly Phe Pro Ser Arg Leu Cys Thr Pro
            560                 565                 570
```

```
Gly Glu Met Val Lys Phe Leu Thr Ala Ile Ile Phe Asn Cys Ser
                575                 580                 585
Ala Gln His Ala Ala Val Asn Ser Gly Gln Met Thr Leu Gly Ala
                590                 595                 600
Trp Met Pro Asn Ala Pro Ser Ser Met Arg Gln Pro Pro Pro Gln
                605                 610                 615
Thr Lys Gly Thr Thr Thr Leu Lys Thr Tyr Leu Asp Thr Leu Pro
                620                 625                 630
Glu Val Asn Ile Ser Cys Asn Asn Leu Leu Leu Phe Trp Leu Val
                635                 640                 645
Ser Gln Glu Pro Lys Asp Gln Arg Pro Leu Gly Thr Tyr Pro Asp
                650                 655                 660
Glu His Phe Thr Glu Glu Ala Pro Arg Arg Ser Ile Ala Ala Phe
                665                 670                 675
Gln Ser Arg Leu Ala Gln Ile Ser Arg Asp Ile Gln Glu Arg Asn
                680                 685                 690
Gln Gly Leu Ala Leu Pro Tyr Thr Tyr Leu Asp Pro Pro Leu Ile
                695                 700                 705
Glu Asn Ser Val Ser Ile
                710
```

<210> 4
<211> 1216
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1321517CD1

<400> 4

```
Met Ala Gly Ala Gln Pro Gly Val His Ala Leu Gln Leu Lys Pro
  1               5                  10                  15
Val Cys Val Ser Asp Ser Leu Lys Lys Gly Thr Lys Phe Val Lys
                20                  25                  30
Trp Asp Asp Asp Ser Thr Ile Val Thr Pro Ile Ile Leu Arg Thr
                35                  40                  45
Asp Pro Gln Gly Phe Phe Phe Tyr Trp Thr Asp Gln Asn Lys Glu
                50                  55                  60
Thr Glu Leu Leu Asp Leu Ser Leu Val Lys Asp Ala Arg Cys Gly
                65                  70                  75
Arg His Ala Lys Ala Pro Lys Asp Pro Lys Leu Arg Glu Leu Leu
                80                  85                  90
Asp Val Gly Asn Ile Gly Arg Leu Glu Gln Arg Met Ile Thr Val
                95                 100                 105
Val Tyr Gly Pro Asp Leu Val Asn Ile Ser His Leu Asn Leu Val
               110                 115                 120
Ala Phe Gln Glu Glu Val Ala Lys Glu Trp Thr Asn Glu Val Phe
               125                 130                 135
Ser Leu Ala Thr Asn Leu Leu Ala Gln Asn Met Ser Arg Asp Ala
               140                 145                 150
Phe Leu Glu Lys Ala Tyr Thr Lys Leu Lys Leu Gln Val Thr Pro
               155                 160                 165
Glu Gly Arg Ile Pro Leu Lys Asn Ile Tyr Arg Leu Phe Ser Ala
               170                 175                 180
Asp Arg Lys Arg Val Glu Ile Ala Leu Glu Ala Cys Ser Leu Pro
               185                 190                 195
Ser Ser Arg Asn Asp Ser Ile Pro Gln Glu Asp Phe Thr Pro Glu
               200                 205                 210
Val Tyr Arg Val Phe Leu Asn Asn Leu Cys Pro Arg Pro Glu Ile
               215                 220                 225
Asp Asn Ile Phe Ser Glu Phe Gly Ala Lys Ser Lys Pro Tyr Leu
               230                 235                 240
```

Thr Val Asp Gln Met Met Asp Phe Ile Asn Leu Lys Gln Arg Asp
245 250 255
Pro Arg Leu Asn Glu Ile Leu Tyr Pro Pro Leu Lys Gln Glu Gln
260 265 270
Val Gln Val Leu Ile Glu Lys Tyr Glu Pro Asn Asn Ser Leu Ala
275 280 285
Arg Lys Gly Gln Ile Ser Val Asp Gly Phe Met Arg Tyr Leu Ser
290 295 300
Gly Glu Glu Asn Gly Val Val Ser Pro Glu Lys Leu Asp Leu Asn
305 310 315
Glu Asp Met Ser Gln Pro Leu Ser His Tyr Phe Ile Asn Ser Ser
320 325 330
His Asn Thr Tyr Leu Thr Ala Gly Gln Leu Ala Gly Asn Ser Ser
335 340 345
Val Glu Met Tyr Arg Gln Val Leu Leu Ser Gly Cys Arg Cys Val
350 355 360
Glu Leu Asp Cys Trp Lys Gly Arg Thr Ala Glu Glu Glu Pro Val
365 370 375
Ile Thr His Gly Phe Thr Met Thr Thr Glu Ile Ser Phe Lys Glu
380 385 390
Val Ile Glu Ala Ile Ala Glu Cys Ala Phe Lys Thr Ser Pro Phe
395 400 405
Pro Ile Leu Leu Ser Phe Glu Asn His Val Asp Ser Pro Lys Gln
410 415 420
Gln Ala Lys Met Ala Glu Tyr Cys Arg Leu Ile Phe Gly Asp Ala
425 430 435
Leu Leu Met Glu Pro Leu Glu Lys Tyr Pro Leu Glu Ser Gly Val
440 445 450
Pro Leu Pro Ser Pro Met Asp Leu Met Tyr Lys Ile Leu Val Lys
455 460 465
Asn Lys Lys Lys Ser His Lys Ser Ser Glu Gly Ser Gly Lys Lys
470 475 480
Lys Leu Ser Glu Gln Ala Ser Asn Thr Tyr Ser Asp Ser Ser Ser
485 490 495
Met Phe Glu Pro Ser Ser Pro Gly Ala Gly Glu Ala Asp Thr Glu
500 505 510
Ser Asp Asp Asp Asp Asp Asp Asp Asp Cys Lys Lys Ser Ser Met
515 520 525
Asp Glu Gly Thr Ala Gly Ser Glu Ala Met Ala Thr Glu Glu Met
530 535 540
Ser Asn Leu Val Asn Tyr Ile Gln Pro Val Lys Phe Glu Ser Phe
545 550 555
Glu Ile Ser Lys Lys Arg Asn Lys Ser Phe Glu Met Ser Ser Phe
560 565 570
Val Glu Thr Lys Gly Leu Glu Gln Leu Thr Lys Ser Pro Val Glu
575 580 585
Phe Val Glu Tyr Asn Lys Met Gln Leu Ser Arg Ile Tyr Pro Lys
590 595 600
Gly Thr Arg Val Asp Ser Ser Asn Tyr Met Pro Gln Leu Phe Trp
605 610 615
Asn Ala Gly Cys Gln Met Val Ala Leu Asn Phe Gln Thr Met Asp
620 625 630
Leu Ala Met Gln Ile Asn Met Gly Met Tyr Glu Tyr Asn Gly Lys
635 640 645
Ser Gly Tyr Arg Leu Lys Pro Glu Phe Met Arg Arg Pro Asp Lys
650 655 660
His Phe Asp Pro Phe Thr Glu Gly Ile Val Asp Gly Ile Val Ala
665 670 675
Asn Thr Leu Ser Val Lys Ile Ile Ser Gly Gln Phe Leu Ser Asp
680 685 690
Lys Lys Val Gly Thr Tyr Val Glu Val Asp Met Phe Gly Leu Pro
695 700 705
Val Asp Thr Arg Arg Lys Ala Phe Lys Thr Lys Thr Ser Gln Gly

710 715 720
Asn Ala Val Asn Pro Val Trp Glu Glu Glu Pro Ile Val Phe Lys
725 730 735
Lys Val Val Leu Pro Thr Leu Ala Cys Leu Arg Ile Ala Val Tyr
740 745 750
Glu Glu Gly Gly Lys Phe Ile Gly His Arg Ile Leu Pro Val Gln
755 760 765
Ala Ile Arg Pro Gly Tyr His Tyr Ile Cys Leu Arg Asn Glu Arg
770 775 780
Asn Gln Pro Leu Thr Leu Pro Ala Val Phe Val Tyr Ile Glu Val
785 790 795
Lys Asp Tyr Val Pro Asp Thr Tyr Ala Asp Val Ile Glu Ala Leu
800 805 810
Ser Asn Pro Ile Arg Tyr Val Asn Leu Met Glu Gln Arg Ala Lys
815 820 825
Gln Leu Ala Ala Leu Thr Leu Glu Asp Glu Glu Glu Val Lys Lys
830 835 840
Glu Ala Asp Pro Gly Glu Thr Pro Ser Glu Ala Pro Ser Glu Ala
845 850 855
Arg Thr Thr Pro Ala Glu Asn Gly Val Asn His Thr Thr Thr Leu
860 865 870
Thr Pro Lys Pro Pro Ser Gln Ala Leu His Ser Gln Pro Ala Pro
875 880 885
Gly Ser Val Lys Ala Pro Ala Lys Thr Glu Asp Leu Ile Gln Ser
890 895 900
Val Leu Thr Glu Val Glu Ala Gln Thr Ile Glu Glu Leu Lys Gln
905 910 915
Gln Lys Ser Phe Val Lys Leu Gln Lys Lys His Tyr Lys Glu Met
920 925 930
Lys Asp Leu Val Lys Arg His His Lys Lys Thr Thr Asp Leu Ile
935 940 945
Lys Glu His Thr Thr Lys Tyr Asn Glu Ile Gln Asn Asp Tyr Leu
950 955 960
Arg Arg Arg Ala Ala Leu Glu Lys Ser Ala Lys Lys Asp Ser Lys
965 970 975
Lys Lys Ser Glu Pro Ser Ser Pro Asp His Gly Ser Ser Thr Ile
980 985 990
Glu Gln Asp Leu Ala Ala Leu Asp Ala Glu Met Thr Gln Lys Leu
995 1000 1005
Ile Asp Leu Lys Asp Lys Gln Gln Gln Gln Leu Leu Asn Leu Arg
1010 1015 1020
Gln Glu Gln Tyr Tyr Ser Glu Lys Tyr Gln Lys Arg Glu His Ile
1025 1030 1035
Lys Leu Leu Ile Gln Lys Leu Thr Asp Val Ala Glu Glu Cys Gln
1040 1045 1050
Asn Asn Gln Leu Lys Lys Leu Lys Glu Ile Cys Glu Lys Glu Lys
1055 1060 1065
Lys Glu Leu Lys Lys Lys Met Asp Lys Lys Arg Gln Glu Lys Ile
1070 1075 1080
Thr Glu Ala Lys Ser Lys Asp Lys Ser Gln Met Glu Glu Glu Lys
1085 1090 1095
Thr Glu Met Ile Arg Ser Tyr Ile Gln Glu Val Val Gln Tyr Ile
1100 1105 1110
Lys Arg Leu Glu Glu Ala Gln Ser Lys Arg Gln Glu Lys Leu Val
1115 1120 1125
Glu Lys His Lys Glu Ile Arg Gln Gln Ile Leu Asp Glu Lys Pro
1130 1135 1140
Lys Leu Gln Val Glu Leu Glu Gln Glu Tyr Gln Asp Lys Phe Lys
1145 1150 1155
Arg Leu Pro Leu Glu Ile Leu Glu Phe Val Gln Glu Ala Met Lys
1160 1165 1170
Gly Lys Ile Ser Glu Asp Ser Asn His Gly Ser Ala Pro Leu Ser
1175 1180 1185

```
Leu Ser Ser Asp Pro Gly Lys Val Asn His Lys Thr Pro Ser Ser
                1190                1195                1200
Glu Glu Leu Gly Gly Asp Ile Pro Gly Lys Glu Phe Asp Thr Pro
                1205                1210                1215
Leu
```

<210> 5
<211> 1239
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472768CD1

<400> 5

```
Met Gly Ala Met Gln Glu Gly Met Gln Met Val Lys Leu Arg Gly
  1               5                  10                  15
Gly Ser Lys Gly Leu Val Arg Phe Tyr Tyr Leu Asp Glu His Arg
             20                  25                  30
Ser Cys Ile Arg Trp Arg Pro Ser Arg Lys Asn Glu Lys Ala Lys
             35                  40                  45
Ile Ser Ile Asp Ser Ile Gln Glu Val Ser Glu Gly Arg Gln Ser
             50                  55                  60
Glu Val Phe Gln Arg Tyr Pro Asp Gly Ser Phe Asp Pro Asn Cys
             65                  70                  75
Cys Phe Ser Ile Tyr His Gly Ser His Arg Glu Ser Leu Asp Leu
             80                  85                  90
Val Ser Thr Ser Ser Glu Val Ala Arg Thr Trp Val Thr Gly Leu
             95                 100                 105
Arg Tyr Leu Met Ala Gly Ile Ser Asp Glu Asp Ser Leu Ala Arg
            110                 115                 120
Arg Gln Arg Thr Arg Asp Gln Tyr Pro Trp Ala Pro Ile Gly Gln
            125                 130                 135
Cys Arg Pro Arg Asp Arg Pro Leu Gly Cys Ser Pro Trp Gly Gly
            140                 145                 150
Leu Ser Phe Ala Gly Ser His Thr Gly Glu Val Ala Gly Gln Arg
            155                 160                 165
Val Glu Trp Leu Lys Gln Thr Phe Asp Glu Ala Asp Lys Asn Gly
            170                 175                 180
Asp Gly Ser Leu Ser Ile Gly Glu Val Leu Gln Leu Leu His Lys
            185                 190                 195
Leu Asn Val Asn Leu Pro Arg Gln Arg Val Lys Gln Met Phe Arg
            200                 205                 210
Val Ala Gly His Ala Trp Leu Glu Gln Gly Lys Leu Ala Cys Ser
            215                 220                 225
Gln Asp Arg Ala Leu Val Glu Val Pro Met Gly Thr Gln Gly Leu
            230                 235                 240
Ala Leu Gln Glu Ala Asp Thr Asp Asp His Gln Gly Thr Leu Gly
            245                 250                 255
Phe Glu Glu Phe Cys Ala Phe Tyr Lys Met Met Ser Thr Arg Arg
            260                 265                 270
Asp Leu Tyr Leu Leu Met Leu Thr Tyr Ser Asn His Lys Asp His
            275                 280                 285
Leu Asp Ala Ala Ser Leu Gln Arg Phe Leu Gln Val Glu Gln Lys
            290                 295                 300
Met Ala Gly Val Thr Leu Glu Ser Cys Gln Asp Ile Ile Glu Gln
            305                 310                 315
Phe Glu Pro Cys Pro Glu Asn Lys Ser Lys Gly Leu Leu Gly Ile
            320                 325                 330
Asp Gly Phe Thr Asn Tyr Thr Arg Ser Pro Ala Gly Asp Ile Phe
            335                 340                 345
```

```
Asn Pro Glu His His His Val His Gln Asp Met Thr Gln Pro Leu
                350                 355                 360
Ser His Tyr Phe Ile Thr Ser Ser His Asn Thr Tyr Leu Val Gly
                365                 370                 375
Asp Gln Leu Met Ser Gln Ser Arg Val Asp Met Tyr Ala Trp Val
                380                 385                 390
Leu Gln Ala Gly Cys Arg Cys Val Glu Val Asp Cys Trp Asp Gly
                395                 400                 405
Pro Asp Gly Glu Pro Ile Val His His Gly Tyr Thr Leu Thr Ser
                410                 415                 420
Lys Ile Leu Phe Lys Asp Val Ile Glu Thr Ile Asn Lys Tyr Ala
                425                 430                 435
Phe Ile Lys Asn Glu Tyr Pro Val Ile Leu Ser Ile Glu Asn His
                440                 445                 450
Cys Ser Val Ile Gln Gln Lys Lys Met Ala Gln Tyr Leu Thr Asp
                455                 460                 465
Ile Leu Gly Asp Lys Leu Asp Leu Ser Ser Val Ser Ser Glu Asp
                470                 475                 480
Ala Thr Thr Leu Pro Ser Pro Gln Met Leu Lys Gly Lys Ile Leu
                485                 490                 495
Val Lys Gly Lys Lys Leu Pro Ala Asn Ile Ser Glu Asp Ala Glu
                500                 505                 510
Glu Gly Glu Val Ser Asp Glu Asp Ser Ala Asp Glu Ile Asp Asp
                515                 520                 525
Asp Cys Lys Leu Leu Asn Gly Asp Ala Ser Thr Asn Arg Lys Arg
                530                 535                 540
Val Glu Asn Thr Ala Lys Arg Lys Leu Asp Ser Leu Ile Lys Glu
                545                 550                 555
Ser Lys Ile Arg Asp Cys Glu Asp Pro Asn Asn Phe Ser Val Ser
                560                 565                 570
Thr Leu Ser Pro Ser Gly Lys Leu Gly Arg Lys Ser Lys Ala Glu
                575                 580                 585
Glu Asp Val Glu Ser Gly Glu Asp Ala Gly Ala Ser Arg Arg Asn
                590                 595                 600
Gly Arg Leu Val Val Gly Ser Phe Ser Arg Arg Lys Lys Lys Gly
                605                 610                 615
Ser Lys Leu Lys Lys Ala Ala Ser Val Glu Glu Gly Asp Glu Gly
                620                 625                 630
Gln Asp Ser Pro Gly Gly Gln Ser Arg Gly Ala Thr Arg Gln Lys
                635                 640                 645
Lys Thr Met Lys Leu Ser Arg Ala Leu Ser Asp Leu Val Lys Tyr
                650                 655                 660
Thr Lys Ser Val Ala Thr His Asp Ile Glu Met Glu Ala Ala Ser
                665                 670                 675
Ser Trp Gln Val Ser Ser Phe Ser Glu Thr Lys Ala His Gln Ile
                680                 685                 690
Leu Gln Gln Lys Pro Ala Gln Tyr Leu Arg Phe Asn Gln Gln Gln
                695                 700                 705
Leu Ser Arg Ile Tyr Pro Ser Ser Tyr Arg Val Asp Ser Ser Asn
                710                 715                 720
Tyr Asn Pro Gln Pro Phe Trp Asn Ala Gly Cys Gln Met Val Ala
                725                 730                 735
Leu Asn Tyr Gln Ser Glu Gly Arg Met Leu Gln Leu Asn Arg Ala
                740                 745                 750
Lys Phe Ser Ala Asn Gly Gly Cys Gly Tyr Val Leu Lys Pro Gly
                755                 760                 765
Cys Met Cys Gln Gly Val Phe Asn Pro Asn Ser Glu Asp Pro Leu
                770                 775                 780
Pro Gly Gln Leu Lys Lys Gln Leu Val Leu Arg Ile Ile Ser Gly
                785                 790                 795
Gln Gln Leu Pro Lys Pro Arg Asp Ser Met Leu Gly Asp Arg Gly
                800                 805                 810
Glu Ile Ile Asp Pro Phe Val Glu Val Glu Ile Ile Gly Leu Pro
```

```
                815                 820                 825
Val Asp Cys Ser Arg Glu Gln Thr Arg Val Val Asp Asp Asn Gly
                830                 835                 840
Phe Asn Pro Thr Trp Glu Glu Thr Leu Val Phe Met Val His Met
                845                 850                 855
Pro Glu Ile Ala Leu Val Arg Phe Leu Val Trp Asp His Asp Pro
                860                 865                 870
Ile Gly Arg Asp Phe Ile Gly Gln Arg Thr Leu Ala Phe Ser Ser
                875                 880                 885
Met Met Pro Gly Tyr Arg His Val Tyr Leu Glu Gly Met Glu Glu
                890                 895                 900
Ala Ser Ile Phe Val His Val Ala Val Ser Asp Ile Ser Gly Lys
                905                 910                 915
Val Lys Gln Ala Leu Gly Leu Lys Gly Leu Phe Leu Arg Gly Pro
                920                 925                 930
Lys Pro Gly Ser Leu Asp Ser His Ala Ala Gly Arg Pro Pro Ala
                935                 940                 945
Arg Pro Ser Val Ser Gln Arg Ile Leu Arg Arg Thr Ala Ser Ala
                950                 955                 960
Pro Thr Lys Ser Gln Lys Pro Gly Arg Arg Gly Phe Pro Glu Leu
                965                 970                 975
Val Leu Gly Thr Arg Asp Thr Gly Ser Lys Gly Val Ala Asp Asp
                980                 985                 990
Val Val Pro Pro Gly Pro Gly Pro Ala Pro Glu Ala Pro Ala Gln
                995                1000                1005
Glu Gly Pro Gly Ser Gly Ser Pro Arg Gly Lys Ala Pro Ala Ala
               1010                1015                1020
Val Ala Glu Lys Ser Pro Val Arg Val Arg Pro Pro Arg Val Leu
               1025                1030                1035
Asp Gly Pro Gly Pro Ala Gly Met Ala Ala Thr Cys Met Lys Cys
               1040                1045                1050
Val Val Gly Ser Cys Ala Gly Val Asn Thr Gly Gly Leu Gln Arg
               1055                1060                1065
Glu Arg Pro Pro Ser Pro Gly Pro Ala Ser Arg Gln Ala Ala Ile
               1070                1075                1080
Arg Gln Gln Pro Arg Ala Arg Ala Asp Ser Leu Gly Ala Pro Cys
               1085                1090                1095
Cys Gly Leu Asp Pro His Ala Ile Pro Gly Arg Ser Arg Glu Ala
               1100                1105                1110
Pro Lys Gly Pro Gly Ala Trp Arg Gln Gly Pro Gly Gly Ser Gly
               1115                1120                1125
Ser Met Ser Ser Asp Ser Ser Ser Pro Asp Ser Pro Gly Ile Pro
               1130                1135                1140
Glu Arg Ser Pro Arg Trp Pro Glu Gly Ala Cys Arg Gln Pro Gly
               1145                1150                1155
Ala Leu Gln Gly Glu Met Ser Ala Leu Phe Ala Gln Lys Leu Glu
               1160                1165                1170
Glu Ile Arg Ser Lys Ser Pro Met Phe Ser Ala Gly Lys Pro Leu
               1175                1180                1185
Leu Pro Cys Val Val Leu Pro His Ala Pro Gly Met Ala Gly Pro
               1190                1195                1200
Gly Ser Pro Ala Ala Ala Ser Ala Trp Thr Val Ser Pro Arg Val
               1205                1210                1215
Leu Val Leu Val Ala Leu Tyr Pro Trp His Cys Leu Arg Gly Thr
               1220                1225                1230
Leu Leu Pro Trp Leu Ala Cys Gly Pro
               1235
```

<210> 6
<211> 4813
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473568CB1

<400> 6
atcagggaga gcctctctga ggaggggata atgaagacaa ggtctgtaga atgaagagga 60
gcctgatgca ccaagggcag tagaaagaga cttctgagca agaggaactg caagtgcaaa 120
ggccctgatg ctagagagag cttgaggggt tgcaggaaaa gagtttttat ggttgtgaac 180
aagaaataga aggaagaaca gagagtggca tgagataaaa tttgaaggat cccagatcgt 240
tcaggacctt ttcggccagg aatcctttgt gtttgaaccc aactgcctct tcaaagtgga 300
tgagtttggc ttctttctga catggagaag tgaaggcaag gaaggacagg tgctagaatg 360
ctccctcatc aacagtattc ggtcgggagc cataccaaag gatcccaaaa tcttggctgc 420
tcttgaagct gttggaaaat cagaaaatga tctggaaggg cggatagttt gtgtctgcag 480
tggcacagat ctagtgaaca ttagttttac ctacatggtg gctgaaaatc cagaagtaac 540
taagcaatgg gtagaaggcc tgagatcaat catacacaac ttcagggcca acaacgtcag 600
tccaatgaca tgcctcaaga aacactggat gaaattggca tttatgacca acacaaatgg 660
taaaattcca gttaggagta ttactagaac atttgcatcg ggaaaaacag aaaaggtgat 720
ctttcaagca ctcaaggagt taggtcttcc cagtggaaag aatgatgaaa ttgagcccac 780
agcattttct tatgaaaagt tctatgaact gacacaaaag atttgtcctc ggacagatat 840
agaagatctt ttcaaaaaaa tcaatggaga caaaactgat tatttaacgg tagaccaatt 900
agtgagcttt ctaaatgaac atcaacgaga tcctcgattg aatgaaattt tatttccatt 960
ttatgatgcc aaaagggcaa tgcagatcat tgagatgtat gaacctgatg aagatttgaa 1020
gaaaaaaggc cttatatcaa gtgatgggtt ttgcagatat ctgatgtcag atgaaaacgc 1080
cccagtcttc ctagatcgtt tagaacttta ccaagaaatg gaccatcctc tggctcacta 1140
cttcatcagt tcttcccata acacttatct cactggcaga cagttcggcg ggaagtcttc 1200
ggtagaaatg tacagacagg ttctcctggc tggttgcaga tgtgttgaac ttgactgctg 1260
ggatggaaaa ggtgaagacc aagaaccaat aataactcat ggaaaagcaa tgtgtacaga 1320
tatccttttt aaggatgtaa ttcaagccat caaggaaact gcatttgtca catcagaata 1380
tcctgtaatt ctctcctttg aaaatcactg cagcaaatat caacagtaca agatgtccaa 1440
atattgcgaa gatctatttg gggatctcct gttgaaacaa gcacttgaat cacatccact 1500
tgaaccaggc agggctttgc catcccccaa tgacctcaaa agaaaaatac tcataaaaaa 1560
caagcggctg aaacctgaag ttgaaaaaaa acagctggaa gctttgagaa gcatgatgga 1620
agctggagaa tctgcctccc cagcaaacat cttagaggac gataatgaag aggagatcga 1680
aagtgctgac caagaggagg aagctcaccc cgaattcaaa tttggaaatg aactttctgc 1740
tgatgacttg ggtcacaagg aagctgttgc aaatagcgtc aagaaggctt cagatgacct 1800
tgaacatgaa aacaacaaaa agggcctggt cactgtagaa gatgagcagg cgtggatggc 1860
atcttataaa tatgtaggtg ctaccactaa tatccatcca tatttgtcca caatgatcaa 1920
ctacgcccag cctgtaaagt ttcaaggttt ccatgtggca gaagaacgca atattcatta 1980
taacatgtct tcttttaatg aatcagtcgg tcttggctac ttgaagacac atgcaattga 2040
atttgtcaat tataacaaac ggcaaatgag tcgcatttac cccaagggag gccgagtcga 2100
ttccagtaat tacatgcctc agattttctg gaacgctggc tgccagatgg tttcactgaa 2160
ctatcaaacc ccagatttag cgatgcaatt gaatcaggga aaatttgagt ataatggatc 2220
gtgcgggtac cttctcaaac cagatttcat gaggcggcct gatcgaacat ttgacccctt 2280
ctctgaaact cctgttgatg gtgttattgc agccacttgc tcagtgcagg ttatatcagg 2340
tcaattctta tcagataaga aaattggcac ctacgtagag gtggatatgt atgggttgcc 2400
cactgacacc atacgtaagg aattccgaac tcgcatggtt atgaataatg gactcaatcc 2460
agtttacaat gaagagtcat ttgtatttcg gaaggtgatc ctgccggacc tggctgtctt 2520
gagaatagct gtgtatgatg ataacaacaa gctgattggc cagaggatcc tcccgcttga 2580
tggcctccaa gccggatatc gacacatttc ccttcgaaat gagggaaata aaccattatc 2640
actaccaaca attttctgca atattgttct taaaacatat gtgcctgatg gatttggaga 2700
tatcgtggat gctttatcag atccaaagaa atttctctca attacagaaa agagagcaga 2760
ccaaatgaga gctatgggca ttgaaactag tgacatagcc gacgtgccca gtgacacttc 2820
caaaaatgac aagaaaggaa aggccaacac cgccaaagca aatgtgaccc ctcagagtag 2880
ctctgagctc agaccaacca ccacggctgc cctggcctct ggtgtggaag ccaagaaagg 2940
tattgaactt atccctcaag taaggataga agacttaaag cagatgaagg cttacttgaa 3000
gcatttaaag aaacagcaga aggagctaaa ttctttaaag aagaaacatg caaaggaaca 3060
cagtaccatg cagaagttac actgcacgca agttgacaaa attgtggcac agtatgacaa 3120
agagaagtcg actcatgaga aaatcctaga gaaggcaatg aagaagaagg ggggaagtaa 3180
ttgtctcgaa atgaaaaaag aaacagaaat caaaattcag acgctgacat cagatcacaa 3240
atctaaggtc aaagagattg tagcacagca cacaaaggaa tggtcagaaa tgatcaatac 3300
ccacagtgct gaggagcaag aaatccgaga cctgcacctc agccagcagt gtgagctgct 3360
gaaaaagcta ctcatcaatg cccacgagca gcaaacccag cagctgaaac tgtcccatga 3420
cagggaaagc aaggaaatgc gagcacacca ggctaagatt tctatggaaa atagcaaagc 3480

catcagccaa gataaatcta tcaagaataa agcagaacgg gaaaggcgag tcagggagtt 3540
aaacagcagc aacactaaaa agtttctgga agaaagaaag agacttgcca tgaagcagtc 3600
caaagaaatg gatcagttga aaaaagtcca gcttgaacat ctagaattcc tagagaaaca 3660
gaatgagcag cttttgaaat cctgtcatgc agtgtcccaa acgcaaggcg aaggagatgc 3720
agcagatggt gaaattggaa gccgagatgg accgcagacc agcaacagta gtatgaaact 3780
ccaaaatgca aactgaagca gcaaacccac aaagcatcaa aagactcact cacaaacttc 3840
tgaacacaaa ctccatggat gaaagctgtt tattttgttt cctttatgtg taaacaagat 3900
gatatctgaa accagagaga cttggaatgt ctgactgact tctatttaac agcttgagta 3960
ttgcatttcc ttggccaaac aaaaatagct acaaatccac aaaaatttac tattccagta 4020
aggcagagtc caaccattga taatacaact taaacatgtt tgctataaaa taccatcaca 4080
agtaaatgag cttggtgtga acaactctcc tttgtgatgc cttaggacat gtttgaactg 4140
cagcaaaaaa caaaaacaaa aaacagtgca ttagcaattt catagcaagt gcatgcacta 4200
ggaaaagaaa actctgtcta caagtttatt agcagaagtg gtggtctgct agacaaataa 4260
ttttgcaaaa tttttctaca tctaagttac ctcatcagta agtgccatgt ctctaccatg 4320
ccatcagagg ctaatttcct gtaaaagttg tggaaattgt tagaacaata gaaaaataga 4380
gcagtgtatg tgtgccaaaa ctcatcatta ctcaaaggag aactgtgtta ggcacattta 4440
agaaagttta catctgacat tgctttatag gaattgtttc tgcagattcc ggatattata 4500
attcacacca taaagattgt gaagtggtta ttggcaaacg tttgtaaatg tgaccatgta 4560
taaagtattt atactcttta attcacactg ttagagagca aaatcatcta agtattgcca 4620
catgacaaga ttagtaaaca ggaatactag aactatgttt gcatgataca caagcaccaa 4680
taaagactaa tccatacaca gttaacctaa tgccaaataa atactggtta aataaatgta 4740
tggcacagaa tataatttga ctatcaagac ttttagcata atgaaaaccc ctccctctat 4800
atatatatgt gta 4813

<210> 7
<211> 2270
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473224CB1

<400> 7
gcagcgtgaa gctggggcct gctccccgca gcctctggag cgcatctcag accttctgag 60
acctatgttg ctggcccccc agaacccgca acatggcaga tggggcaaag gccaacccca 120
aagggttcaa aaagaaggtg ctggatagat gcttctctgg gtggaggggc ccacgcttcg 180
gggcctcctg tccttcaaga acctccaggt ctagcctggg tatgaagaag ttcttcaccg 240
tggccatcct tgctggcagc gttctgtcca cagctcacgg cagcctgctc aacctgaagg 300
ccatggtgga ggccgtcaca gggaggagcg ccatcctgtc cttcgtgggc tacggttgct 360
actgtgggct ggggggccgt ggccagccca aggatgaggt ggactggtgc tgccacgccc 420
acgactgctg ctaccaggaa ctctttgacc aaggctgtca cccctatgtg gaccactatg 480
atcacaccat cgagaacaac actgagatag tctgcagtga cctcaacaag acagagtgtg 540
acaagcagac atgcatgtgt gacaagaaca tggttctgtg cctcatgaac cagacgtacc 600
gagaggagta ccgtggcttc ctcaatgtct actgccaggg ccccacgccc aactgcagca 660
tctatgaacc gccccctgag gaggtcacct gcagtcacca atccccagcg ccccccgccc 720
ctccctagag cctctgaggt ttgagagaga gagcgggagg agggtctggc ttggggacca 780
gacgaggtgc agggagggta ggagccaggc caggagcctg agggttgctg gttgcctcct 840
ccctggagct ctccagtgag ggctcagctc tcagaggact caggaaggcc tgggtcctga 900
ctcccccagc ccagccccag gcatgggtgc ctcctgctgc tggttctgga ctgggtggga 960
ggcacggagc ttataggggt ctctcctgag ggtggccggg gagacctgag agagaggagg 1020
aggggcctct gagtggggcc tctgttgctg gcgccagttt aactccccgg agccttagaa 1080
agtctgagcc ttagaaagtc tgggcctgag catccaggcc cagagctgga tgcatcctcg 1140
gcccaagatc acaggaaggc agattgctgg tcacaaagca tgggttccgg gagcccctca 1200
gctgatccca caggatggcc tggggtggtg gctactttgg gcttgaagct ctctagagcc 1260
ccattcacag agcaggtgcc ccaccccagc cctcactgag gtggccatga gtccaggttc 1320
agaccctagg ggatgtgcag tgctgagtgt cccagccctg aagtctagcc tcagagcagg 1380
ctcctgggca ctgacagagc cacacaccca ggtacacccc ccaggctaat atggggacac 1440
acaaaaatat ggccttcaac acctacaagt cttctctgtg tcccccatcg acccatgcac 1500
atcaccccaa ccctgctctg ccacaccacg tgggtctgag gagcactcct acccccggc 1560
tccgggtccc tgacagacac tgccctcctg gcctgcacct ggaatggcag ccagaacact 1620
ggaggggggac ccacaccctc tctccctggg ctgtggatgg gtgttgcaag tcccatccct 1680
cctggcctgc acctggaaca gcagccagaa caccggaggg ggacccacac cctctctccc 1740

tgggctgtgg gcaggcgtca caagtcccat tggtggggaa gaggctgagg gctgagaccc 1800
acctcagtga agagggaaac gtaagcaggg agagcatcaa ggcgggaggc agcctggaaa 1860
cttctagaag agggtcaagg taggaggcta catggaagct tctagaaggc agtaaggaac 1920
agggtggtga aggagagggg agaagatggg agcatggaga gacgaggtga gtgacagcca 1980
ccagcaggga cgggcctcca ggcaacacag aggagctggg ggccagggcc ttacggggct 2040
gggttcctgt gtggggcgtg gagctggggc tgaagtggag gggacggccc tgcggccccc 2100
agcactgttg tttctccagg tgcaccttac cttcctcctt acctgcccga gcctcaactc 2160
agggtcagca gcctccacgg agccagcccc acctgcccga aaaccactga agccaccctc 2220
cgcccgacgt caacggggat ccactagttt agagcgcggc ccccgcgtgc 2270

<210> 8
<211> 2136
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473234CB1

<400> 8
atggcagtgt accgcctgtg tgtgaccact ggtccctacc tgagggccgg cacactggac 60
aacatctctg tcacactggt gggcacgtgt ggtgaaagcc ccaagcagcg gctagatcga 120
atgggcaggg acttcgcccc tggatcggta cagaagtaca aggtgcgttg cacagcggag 180
ctgggtgagc tcttgctgct gcgtgtacac aaggagcgct acgctttctt ccgcaaggac 240
tcttggtact gtagccgcat ctgtgtcacc gaaccggatg gtagtgtatc ccacttcccc 300
tgctatcagt ggattgaagg ctactgcacc gtggagctga ggccaggaac agcaagaact 360
atttgtcagg actctcttcc cctcctcctg gatcacagga cacgggagct ccgggcccga 420
caagaatgct accgctggaa gatctatgcc cctggcttcc cctgcatggt agacgtcaac 480
agctttcagg agatggagtc agacaagaaa tttgccttga caaagacgac aacttgtgta 540
gaccagggtg acagcagtgg gaatcggtac ctgcccggct tccccatgaa aattgacatc 600
ccatccctga tgtacatgga gcccaatgtt cgatactcag ccaccaagac gatctcgctg 660
ctcttcaatg ccatccctgc gtccttggga atgaagcttc gagggctgtt ggatcgcaag 720
ggctcctgga agaagctgga tgacatgcag aacatcttct ggtgccataa gaccttcacg 780
acaaagtatg tcacagagca ctggtgtgaa gatcacttct ttgggtacca gtacctgaat 840
ggtgtcaatc ccgtcatgct ccactgcatc tctagcttgc ccagcaagct gcctgtcacc 900
aatgacatgg tggccccctt gctgggacag gacacatgcc tgcagacaga gctagagagg 960
gggaacatct tcctagcgga ctactggatc ctggcggagg cccccaccca ctgcctaaac 1020
ggccgccagc agtacgtggc cgccccactg tgcctgctgt ggctcagccc ccaggggcg 1080
ctggtgccct tggccatcca gctcagccag acccccgggc ctgacagccc catcttcctg 1140
cccactgact ccgaatggga ctggctgctg gccaagacgt gggtgcgcaa ctctgagttc 1200
ctggtgcacg aaaacaacac gcactttctg tgcacgcatt tgctgtgcga ggccttcgcc 1260
atggccacgc tgcgccagct gccgctctgc caccccatct acaagctcct actcccccac 1320
actcgataca cgctgcaggt gaacaccatc gcgagggcca cgctgctcaa ccccgagggc 1380
ctcgtggacc aggtcacgtc catcgggagg caaggcctca tctacctcat gagcacgggc 1440
ctggcccact tcacctacac caatttctgc cttccggaca gcctgcgggc ccgcggcgtc 1500
ctggctatcc ccaactacca ctaccgagac gacggcctga agatctgggc ggccattgag 1560
agctttgtct cagaaatcgt gggctactat tatcccagtg acgcatctgt gcagcaggat 1620
tcggagctgc aggcctggac tggcgagatt tttgctcagg cgttcctggg ccgggaaagc 1680
tcaggtttcc caagccggct gtgcacccca ggagagatgg tgaagttcct cactgcaatc 1740
atcttcaatt gctctgccca gcacgctgct gtcaacagtg ggcagatgac tttgggggcc 1800
tggatgccca atgctccatc atccatgagg cagcccccac cccagaccaa ggggaccacc 1860
accctgaaga cttacctaga caccctccct gaagtgaaca tcagctgtaa caacctcctc 1920
ctcttctggt tggttagcca agaacccaag gaccagaggc ccctgggcac ctacccagat 1980
gagcacttca cagaggaggc cccgaggcgg agcatcgccg ccttccagag ccgcctggcc 2040
cagatctcaa gggacatcca ggagcggaac cagggtctgg cactgcccta cacctacctg 2100
gaccctcccc tcattgagaa cagcgtctcc atctaa 2136

<210> 9
<211> 3700
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 1321517CB1

<400> 9
ccccagtccc tgccgcgctc gcccgggccg cccggagccc agatgagccc agatggccgg 60
ggctcaaccc ggagtgcacg ccttgcaact caagcccgtg tgcgtgtccg acagcctcaa 120
gaagggcacc aaattcgtca agtgggatga tgactcaact attgttactc caattatttt 180
gaggactgac cctcagggat ttttctttta ctggacagat caaaacaagg agacagagct 240
actggatctc agccttgtca aagatgccag atgtgggaga cacgccaaag ctcccaagga 300
ccccaaatta cgtgaacttt tggatgtggg gaacatcggg cgcctggagc agcgcatgat 360
cacagtggtg tatgggcctg acctcgtgaa catctcccat ttgaatctcg tggctttcca 420
agaagaagtg gccaaggaat ggacaaatga ggttttcagt ttggcaacaa acctgctggc 480
ccaaaacatg tccagggatg catttctgga aaaagcctat actaaactta agctgcaagt 540
cactccagaa gggcgtattc ctctcaaaaa catatatcgc ttgttttcag cagatcggaa 600
gcgagttgaa attgctttag aggcttgtag tcttccatct tcaaggaatg attcaatacc 660
tcaagaagat ttcactccag aagtgtacag agttttcctc aacaaccttt gccctcgacc 720
tgaaattgat aacatctttt cagaatttgg tgcaaaaagc aaaccatatc ttaccgttga 780
tcagatgatg gattttatca accttaagca gcgagatcct cggcttaatg aaatacttta 840
tccacctcta aaacaagagc aagtccaagt attgattgag aagtatgaac ccaacaacag 900
cctcgccaga aaaggacaaa tatcagtgga tgggttcatg cgctatctga gtggagaaga 960
aaacggagtc gtttcacctg agaaactgga tttgaatgaa gacatgtctc agccccttc 1020
tcactatttc attaattcct cgcacaacac ctacctcaca gctggccaac tggctggaaa 1080
ctcctctgtt gagatgtatc gccaagtgct cctgtctggt tgtcgctgtg tggagctgga 1140
ctgctggaag ggacggactg cagaagagga acctgtcatc acccatggct tcaccatgac 1200
aactgaaata tctttcaagg aagtgataga agcaattgcg gagtgtgcat ttaagacttc 1260
acctttcca attctccttt cgtttgagaa ccatgtggat tccccaaagc agcaagccaa 1320
gatggcggag tactgccgac tgatctttgg ggatgccctt ctcatggagc ccctggaaaa 1380
atatccactg gaatctggag ttcctcttcc aagccctatg gatttaatgt ataaaatttt 1440
ggtgaaaaat aagaagaaat cacacaagtc atcagaagga agcggcaaaa agaagctctc 1500
agaacaagcc tccaacacct acagtgactc ctccagcatg ttcgagccct catccccagg 1560
agccggagaa gctgatacgg aaagtgacga cgacgatgat gatgatgact gtaaaaaatc 1620
ttcaatggat gaggggactg ctggaagtga ggctatggcc acagaagaaa tgtctaatct 1680
ggtgaactat attcagccag tcaagtttga gtcatttgaa atttcaaaaa aaagaaataa 1740
aagttttgaa atgtcttcct tcgtggaaac caaaggactt gaacaactca ccaagtctcc 1800
agtggaattt gtagaatata acaaaatgca gcttagcagg atatatccaa aaggaacacg 1860
tgtggattca tccaactata tgcctcagct cttctggaat gcaggttgtc agatggtggc 1920
acttaatttc cagacaatgg acctggctat gcaaataaat atggggatgt atgaatacaa 1980
cgggaagagt ggctacagat tgaagccaga gttcatgagg aggcctgaca agcattttga 2040
tccatttact gaaggcatcg tagatgggat agtggcaaac actttgtctg ttaagattat 2100
ttcaggtcag tttctttctg ataagaaagt tgggacttac gtggaagtag atatgtttgg 2160
tttgcctgtg gatacaagga ggaaggcatt taagaccaaa acatcccaag gaaatgctgt 2220
gaatcctgtc tgggaagaag aacctattgt gttcaaaaag gtggttcttc ctactctggc 2280
ctgtttgaga atagcagttt atgaagaagg aggtaaattc attggccacc gtatcttgcc 2340
agtgcaagcc attcggccag gctatcacta tatctgtcta aggaatgaaa ggaaccagcc 2400
tctgacgctg cctgctgtct ttgtctacat agaagtgaaa gactatgtgc cagacacata 2460
tgcagatgtc atcgaagctt tatcaaaccc aatccgatat gtgaacctga tggaacagag 2520
agctaagcaa ttggctgctt tgacactgga agatgaagaa gaagtaaaga aagaggctga 2580
tcctggagaa acaccatcag aggctccaag tgaagcgaga acgactccag cagaaaatgg 2640
ggtgaatcac actacaaccc tgacacccaa gccaccctcc caggctctcc acagccagcc 2700
agctccaggt tctgtaaagg cacctgccaa aacagaagat cttattcaga gtgtcttaac 2760
agaagtggaa gcacagacca tcgaagaact aaagcaacag aaatcgtttg tgaaacttca 2820
aaagaaacac tacaaagaaa tgaaagacct ggttaagaga caccacaaga aaaccactga 2880
ccttatcaaa gaacacacta ccaagtataa tgaaattcag aatgactact tgagaaggag 2940
agccgctttg gaaaagtccg ccaaaaagga cagtaagaaa aaatcggaac ccagcagccc 3000
tgatcatggt tcatcaacga ttgagcaaga cctcgctgct ctggatgctg aaatgaccca 3060
aaagttaata gacttgaagg acaaacaaca gcagcagctg cttaatcttc ggcaagaaca 3120
gtattatagt gaaaaatacc agaagcgaga acatattaaa ctgcttattc aaaagttgac 3180
ggatgtcgca gaagagtgtc agaacaatca gttaaagaag ctcaaagaaa tctgtgagaa 3240
agaaaagaaa gaattaaaga agaaaatgga taaaaagagg caggagaaga taacagaagc 3300
taaatccaaa gacaaaagtc agatggaaga ggagaagaca gagatgatcc ggtcatatat 3360
ccaggaagtg gtgcagtata tcaagaggct agaagaagcg caaagtaaac ggcaagaaaa 3420
actcgtagag aaacacaagg aaatacgtca gcagatcctg gatgaaaagc ccaagctgca 3480
ggtggagctg gagcaagaat accaagacaa attcaaaaga ctgcccctcg agattttgga 3540

attcgtgcag gaagccatga aaggaaagat cagtgaagac agcaatcacg gttctgcccc 3600
tctctccctg tcctcagacc ctggaaaagt gaaccacaag actccctcca gtgaggagct 3660
gggaggagac atcccaggaa aagaatttga tactcctctg 3700

<210> 10
<211> 4237
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472768CB1

<400> 10
gagtggccgt gcctccgctg acggccgtgt ctctcctgcg aggagtggct gtgcctccgc 60
tgacagccgt gtctctcccg tgtccagtgg agcggtgcat gggtgccatg caagagggga 120
tgcagatggt gaagctgcgt ggcggctcca agggcctggt ccgcttctac tacctggacg 180
agcaccgctc ctgcatccgc tggaggccct cacgcaagaa cgagaaggcc aagatctcca 240
tcgactccat ccaggaggtg agtgaggggc ggcagtcgga ggtcttccag cgctaccctg 300
acggcagctt cgaccccaac tgctgcttca gcatctacca cggcagccac cgcgagtcgc 360
tggacctggt ctccaccagc agcgaggtgg cgcgcacctg ggtcactggc ctgcgctacc 420
tcatggccgg catcagcgac gaggacagcc tggctcgccg ccagcgcacc agggaccaat 480
atccttgggc acctatcggg caatgcagac ccagggaccg gccccttggc tgctcaccct 540
ggggggggcct gtcctttgcc gggtcacaca ctggggaagt ggccggccag agggtggagt 600
ggctgaagca gacgtttgac gaggccgaca agaacgggga tggcagcctg agcattggcg 660
aggtcctgca gctgctgcac aagctcaacg tgaacctgcc ccggcagagg gtgaagcaga 720
tgttcagggt ggctggtcat gcctggctgg agcaagggaa gctggcctgc tcacaggaca 780
gggccctggt cgaggtgcca atggggaccc aaggccttgc attgcaggaa gcggacacgg 840
atgaccacca agggacgctg ggttttgaag agttctgtgc cttctacaag atgatgtcca 900
cccgccggga cctctacctg ctcatgctga cctacagcaa ccacaaggac cacctggatg 960
ccgccagcct gcagcgcttc ctgcaggtgg agcagaagat ggcgggtgtg accctcgaga 1020
gctgccagga catcatcgag cagtttgagc catgcccaga aaacaagagt aaggggctgc 1080
tgggcattga tggcttcacc aactacacca ggagccctgc tggtgacatc ttcaaccctg 1140
agcaccacca tgtgcaccag gacatgacgc agccgctgag ccactacttc atcacctcgt 1200
cccacaacac ctacctcgtg ggtgaccagc tcatgtccca gtcacgggtg gacatgtatg 1260
cttgggtcct gcaggctggc tgccgctgcg tggaggtgga ctgctgggat gggcccgacg 1320
gggagcccat tgtgcaccat ggctacactc tgacttccaa gatcctcttc aaagacgtca 1380
ttgaaaccat caacaaatat gccttcatca agaatgagta cccagtgatc ctgtccatcg 1440
aaaaccactg cagtgtcatc cagcagaaga aaatggccca gtatctgact gacatccttg 1500
gggacaagct ggacctgtca tcagtgagca gtgaagatgc caccacactc ccctctccac 1560
agatgctcaa gggcaagatc ctcgtgaagg ggaagaagct cccagccaac atcagcgagg 1620
atgcggagga aggcgaggtg tctgatgagg acagtgctga tgagattgac gatgactgca 1680
agctcctcaa tggggatgca tccaccaatc gaaagcgtgt agaaaacact gctaagagga 1740
aactggattc cctcatcaaa gagtcgaaga ttcgggactg tgaggacccc aacaacttct 1800
ccgtctccac actgtcccca tctggaaagc tcggacgcaa gagcaaggct gaagaggacg 1860
tggagtctgg ggaggatgcc ggggccagca gacgcaatgg ccgcctcgtc gtgggaagct 1920
tctccaggcg caagaagaag ggcagcaagc tgaagaaggc ggccagcgtg gaggagggag 1980
atgagggtca ggactccccg ggaggccaga gccgaggggc gacccggcag aagaagacca 2040
tgaagctgtc ccgggccctc tctgacctgg tgaagtacac caagtccgtg gccacccacg 2100
acatagagat ggaggcggcg tccagctggc aggtgtcgtc cttcagcgag accaaggccc 2160
accagattct gcagcagaag ccggcgcagt acctacgctt caaccagcag cagctctccc 2220
gcatctaccc ctcctcctac cgtgtggact ccagcaacta caacccgcag cccttctgga 2280
acgccggctg ccaaatggtt gccctgaact accagtcaga ggggcggatg ctgcagctga 2340
accgagccaa gttcagcgcc aacggtggct gcggctacgt actcaagcct gggtgcatgt 2400
gccagggcgt gttcaacccc aactcggagg acccccctgcc cgggcagctc aagaagcagc 2460
tggtgctccg gatcatcagt ggccagcagc ttcccaagcc gcgcgactcc atgctggggg 2520
accgtgggga gatcatcgac ccctttgtgg aggtggagat cattgggctc cctgtggact 2580
gcagcaggga gcagacccgc gtggtggacg acaacgggtt caaccccacc tgggaggaga 2640
ccctggtttt catggtgcac atgccggaga tcgcgctggt ccgcttcctc gtctgggacc 2700
acgatcccat cgggcgtgac ttcattggcc agaggacgct ggccttcagc agcatgatgc 2760
caggctacag acacgtgtac ctggagggga tggaagaggc ctccatcttc gtgcatgtgg 2820
ctgtcagtga catcagcggt aaggtcaagc aggctctggg cctaaaaggc ctcttcctcc 2880
gaggcccaaa gcccggctcg ctggacagtc atgctgctgg gcggcccccg gcccggccct 2940

```
ccgttagcca gcggatcctg cggcgcacgg ccagcgcccc gaccaagagc cagaagccgg 3000
gccgcagggg cttcccggag ctggtcctgg gtacacggga cacaggctcc aaggggggtgg 3060
cagacgatgt ggtgcccccc gggcccggac ctgctccgga agccccagcc caggaggggc 3120
ccggcagcgg cagcccccga ggtaaggcgc cagctgcggt ggcagagaag agccctgtgc 3180
gagtgcggcc cccgcgtgtc ctggacggcc ccgggcctgc tgggatggcc gccacatgca 3240
tgaagtgtgt ggtgggatcc tgcgccggcg tgaacaccgg gggcctgcag agggagcggc 3300
cacccagccc ggggcctgca agcaggcagg cagccattcg ccagcagccc cgggcccggg 3360
ctgactcact gggggccccc tgctgtggcc tggaccctca cgctatcccg gggagaagca 3420
gagaggcccc caagggtcct ggggcctgga ggcagggtcc aggcggtagc ggctccatgt 3480
cctcggactc cagcagccca gacagcccgg gcatccccga aaggtccccc cgctggcctg 3540
agggtgcctg caggcaaccg ggggccctgc agggagagat gagtgccttg tttgcccaaa 3600
agctggagga gatcaggagt aaatccccca tgttctccgc cggtaagccc ctcttgccct 3660
gcgtggtcct cccgcacgcc cctggcatgg ctgggcctgg gtcacctgct gctgcttctg 3720
cgtggacggt gtcgcctcgt gtgctcgtgc tcgtggctct gtatccgtgg cactgtctcc 3780
gtggcactct gctcccttgg cttgcctgtg gcccatagcc ccagccctcc tgtctgagct 3840
tgaggccctg ggacttgggt ggagctggtt tgaggcccga caggctggga agaaccagct 3900
gctcttgctg agggtctggg gccgggactg tggcctgaca tgctgggccc ctccggctgg 3960
gcgcttcccc aaactcacct cctgggcggc tggcgacctg catggccct gatgcctttc 4020
ctgggactgg gggccacgta ccatcccatt cccacctccc tctagggcag gctccagggg 4080
tccctactgg gaagtctgat gtgggcaggt agtgcagctg ctgggcgtct cctgcgcccc 4140
tgggacgcct ggagcctgct gagtgctgcg tggagtagat tccctgggcc ccagggcttc 4200
gctgctttgg gctgaagcac cccactagaa gggtgct                          4237
```